

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**Avaliação do Impacto do Horizonte Temporal nas Estimativas de EAD de  
Linhas de Crédito com Limites Explícitos a PME**

**VERSÃO PÚBLICA**

**Mestrado em Matemática Aplicada à Economia e Gestão**

Patrícia Alexandra Henriques Silva

Dissertação orientada por:  
Teresa Alpuim  
Vasco Corrêa de Oliveira

Abril, 2016

## **AGRADECIMENTOS**

Esta tese não teria sido possível sem a colaboração de várias pessoas, as quais gostaria de nomear.

Em primeiro lugar, gostaria de agradecer à minha orientadora, a Professora Doutora Teresa Alpuim pela constante disponibilidade, dedicação e orientação ao longo do desenvolvimento deste trabalho. Um agradecimento especial pelas largas horas empenhadas, em conjunto, na resolução dos desafios conceptuais que iam surgindo e pelo seu entusiasmo constante que me motivou a dar o melhor de mim.

Agradeço também ao meu coorientador, Doutor Vasco Corrêa de Oliveira, por todo o apoio na estruturação e desenvolvimento deste trabalho. Pela tempo que me dedicou no esclarecimento de dúvidas e questões que iam surgindo, propondo ideias de melhoramento e sugerindo textos regulamentares importantes para a compreensão do enquadramento regulamentar dos modelos aqui apresentados.

Agradeço também:

- \* Ao Sérgio Naito pelo apoio técnico que me deu na fase de desenvolvimento dos modelos, pela visão que me transmitiu sobre a realidade e os desafios das IF e por me inspirar ao conhecimento que a informação estatística proporciona;
- \* Ao meu colega Jorge Borges pela amizade e ensinamentos proporcionados ao longo deste mestrado;
- \* Às minhas amigas: Paulina Amaral, Tamara Silva, Sandra Andrade, Luísa Antunes, Fabiana Marques e Rute Fonseca por me proporcionarem a confiança e a motivação para abraçar com entusiasmo este desafio;
- \* Aos meus irmãos, João Silva e Sílvia Silva, por serem incondicionais em todos os momentos, por me motivarem nas fases mais difíceis e pela confiança que depositam em mim.
- \* Aos meus pais, pelos ensinamentos que me transmitiram, pela força que sempre me deram e por serem o exemplo de que a dignidade do ser humano nasce da sinceridade do seu trabalho.

## ABSTRACT

Lines of Credit with explicit limits are a major source of funding for SMEs, particularly on giving them support to their cash management. The restricted financial information these companies disclose hinders their access to capital markets, forcing them to rely heavily on bank financing. Additionally, due to the bigger risks they represent to financial institution, their lines of credit are penalized on less availability of the undrawn amount and on higher costs (that is, higher interest rates). In fact, at the time of granting a credit line, the financial institution incurs not only in the risk of default, but also in the risk that the debt at the time of default will be greater than the amount currently used (that is, that the exposure at the time of default, which is known as EAD, will be higher than the amount used on a date before). This is due to the flexibility of this type of financing.

In the European regulatory context, this credit risk should be covered on the one hand, by provisions, which represent a reserve for expected losses reflected on the institution's income statement and, on the other hand, by a capital reserve for unexpected losses in order to ensure the institution's solvency. In the IRB approach established in the European regulatory framework, a financial institution can develop and use a credit risk model to predict increases in the undrawn amount, that is, to predict the EAD. Although this regulatory framework establishes some specific requirements on the development of models (including the EAD measure, the minimum sample period and the need to reflect on estimates the effects of an economic recession), is permissive as to estimation techniques, methods for collecting the observed CCFs and the treatment of outliers (which are problematic in these models).

The EAD measure established in this regulation is called the conversion factor (CCF) and represents the percentage of the amount not yet used that is considered it will be used when the company defaults. Thus, the EAD is the sum of the amount currently used and a percentage (CCF) of the amount not used that the institution predicts it will be at risk when the company defaults.

EAD/CCF models are developed based on this EAD measure and the sample used is collected from a group of defaulted companies. Based on the CCF definition, it is assumed that the greater the distance between the reference date and the default date the higher will be the CCF, as the greater will be the difference in the amount used between the two dates, once the company will tend to drawn more heavily the closer it is to default. In this

sense, the sample method for collecting observed CCFs underlies the need for deciding which reference date to use in order to obtain a set of observations.

This decision taken by financial institutions does not have any regulatory restriction. Generally, there are two main methods used, the variable time horizon (which measures the increases in the undrawn amount along the 12 months before default) and the fixed time horizon (which observes the increase in the undrawn amount exactly the 12th month before default). The second tends to be more conservative, with higher CCF averages, but only uses one month observation for each company. The first method allows the financial institution to collect more information about the company, but requires greater capability to store information about customers and gives more dispersion of observed CCFs.

In this context, this report aims to compare the impact on the EAD estimates at a regulatory level of this two methodologies (the variable time horizon and the fixed time horizon), considering different approaches in terms of estimation techniques and options for the treatment of extreme values. Respecting to estimation techniques, it was used the most common techniques among financial institutions, namely decision trees and multiple linear regression.

Additionally, this report proposes to test which variables in this portfolio are best at explaining increases on the undrawn amount until the time of default. Unlike other studies, data contained historical relationship between the financial institution and the company, which allowed us to evaluate their relevance on predicting additional increases on lines of credit. Typically this information is not available in public databases, thus unlike previous studies, the models presented here may have a better approximation on the adequacy, at a regulatory level, of the most common models used by financial institutions.

In conclusion, this report aims to conclude for the SME portfolio the impact of the time horizon used to obtain a set of observed CCFs on the estimated EAD, looking to identify which credit risk model has the best performance and also can ensure a smaller risk of underestimation, thus guarantying a higher level of conservatism on a regulatory context.

**Keywords:** EAD/CCF, Variable Time Horizon, Fixed Time Horizon, Estimation Techniques, Regulatory Framework

## RESUMO

As linhas de crédito com limites explícitos são uma das principais fontes de financiamento das PME, sobretudo no suporte financeiro que garantem à gestão de tesouraria. A reduzida informação financeira que estas empresas divulgam dificulta o seu acesso aos mercados de capitais, forçando-as a depender quase exclusivamente do financiamento bancário. No acesso ao crédito bancário são objeto ainda de maiores penalizações, quer na disponibilidade das linhas quer nos custos de financiamento (i.e. taxas de juro mais elevadas), devido ao elevado risco que representam para as IF. De facto, no momento de concessão de uma linha de crédito, a IF incorre não apenas no risco de incumprimento por parte da empresa, mas também no risco de que a dívida seja superior ao montante atualmente utilizado pela empresa (i.e. que a exposição no momento de incumprimento, designada por “EAD – *Exposure At Default*”, seja superior ao montante utilizado numa data antes do incumprimento) devido à flexibilidade de utilização que estas linhas permitem.

No contexto regulamentar europeu, este risco de crédito deve ser coberto por um lado, por provisões, as quais constituem uma reserva nos resultados da IF para perdas esperadas, e por outro por uma reserva de fundos próprios que garantam a solvabilidade da IF caso se verifiquem perdas não esperadas. No método IRB, estabelecido no quadro regulamentar europeu, sobre o qual se devem reger as IF pertencentes aos estados-membros da UE, as IF podem desenvolver e utilizar modelos de risco de crédito sobre aumentos da exposição, isto é, sobre o parâmetro EAD. Este quadro regulamentar, embora estabeleça alguns requisitos específicos para o desenvolvimento dos modelos (nomeadamente a medida do EAD a utilizar, o período mínimo da amostra e a necessidade de refletir nas estimativas os efeitos de um período de recessão da economia), é permissivo quanto às técnicas de estimação, aos métodos de recolha das observações e ao tratamento de valores extremos (problemáticos nestes modelos).

A medida necessária à estimativa do EAD, estabelecida nesta regulamentação, designa-se por fator de conversão (CCF) e representa a percentagem do montante ainda não utilizado que se considera que será utilizado quando a empresa incumprir. Assim, o EAD é a soma, num determinado momento de referência, do montante utilizado da linha com a percentagem (o CCF) do montante ainda não utilizado que se assume será utilizado até ao momento em que a empresa entrar em incumprimento. Com base nesta medida são desenvolvidos os modelos de risco de crédito do EAD/CCF, sendo a amostra utilizada

para a estimativa, recolhida de um conjunto de empresas com linhas de crédito que incumpriram. Pela definição do CCF, pressupõe-se que quanto maior for a distância temporal entre a data de referência e a data de incumprimento maior será o CCF, isto é, maior será o diferencial no montante utilizado entre as duas datas, pois a empresa tenderá a utilizar o máximo disponível quanto mais próxima estiver de incumprir. Neste sentido, o método de recolha da amostra tem impacto no CCF, pois pressupõe a definição das datas de referência a utilizar para obter o conjunto dos CCF observados sobre os quais se desenvolvem os modelos estatísticos.

Esta decisão a tomar pelas IF não tem qualquer limitação regulamentar. Em geral, existem 2 principais métodos utilizados pelas IF, o horizonte temporal variável (que observa os acréscimos no montante utilizado durante 12 meses antes do incumprimento, isto é, com 12 datas de referência) e o horizonte temporal fixo (que observa o acréscimo exatamente no 12.º mês anterior ao incumprimento). O segundo método tende a ser mais conservador, com médias de CCF mais elevadas, mas apenas utiliza um mês de observação por empresa. O primeiro método permite recolher mais informação sobre a empresa, mas exige maior capacidade por parte das IF em guardar informação sobre os seus clientes e apresenta maior dispersão nos CCF observados.

Neste contexto, o presente trabalho visa comparar o impacto no contexto regulamentar de ambas as metodologias de recolha nas estimativas de EAD/CCF, considerando diferentes abordagens de técnicas de estimação e de opções de tratamento de valores extremos. Nas técnicas de estimação procurou-se testar para este portefólio a *performance* das técnicas mais comuns entre as IF, nomeadamente as árvores de decisão e a regressão linear múltipla.

Adicionalmente, procurou-se ainda testar as variáveis que neste portefólio melhor explicam acréscimos adicionais de utilização das linhas de crédito até ao momento do incumprimento ou, por outras palavras, as variáveis explicativas do parâmetro CCF. Contrariamente a alguns outros estudos realizados sobre a estimação do EAD/CCF, a base de dados utilizada continha o histórico da relação entre a IF e a empresa, permitindo inferir sobre a relevância destas variáveis na avaliação do risco de acréscimos no nível de utilização das linhas. Esta informação tipicamente não está disponível em bases de dados disponibilizadas por bancos centrais, pelo que, ao contrário de outros estudos realizados, os modelos aqui apresentados permitem avaliar com melhor aproximação à realidade a

adequabilidade, no contexto regulamentar, dos modelos de estimação do EAD/CCF mais utilizados pelas IF.

Em suma, no presente trabalho propõe-se concluir para o portefólio de PME sobre o impacto do horizonte temporal utilizado no EAD estimado, procurando-se identificar os modelos de risco de crédito mais eficientes, isto é, que garantam um bom desempenho e apresentem menor risco de subestimação do EAD, isto é, maior grau de conservadorismo no contexto regulamentar.

**Palavras-chave:** EAD/CCF, Horizonte temporal variável, Horizonte Temporal Fixo, Técnicas de Estimação, Contexto regulamentar

## SIMBOLOGIA E NOTAÇÕES

IF – instituições financeiras são empresas cujo objeto da atividade consiste em receber do público depósitos ou outros fundos reembolsáveis para aplicarem, por sua conta, na concessão de crédito. No grupo das instituições financeiras destacam-se os bancos, mas dele também fazem parte as caixas económicas, a Caixa Central de Crédito Agrícola Mútuo e as caixas de crédito agrícola mútuo. As IF em Portugal adotam as normas regulamentares europeias e são supervisionadas pela autoridade competente, o Banco de Portugal.

PME – Pequena e média empresa

PD – *Probability of Default* – parâmetro regulamentar estabelecido nos Acordos de Basileia, que corresponde à probabilidade de uma empresa/particular entrar em incumprimento dentro do horizonte temporal de um ano.

LGD – *Loss Given Default* – parâmetro regulamentar estabelecido nos Acordos de Basileia, que corresponde à percentagem da exposição na data de incumprimento (EAD) que não será recuperada.

EAD – *Exposure At Default* – parâmetro regulamentar estabelecido nos Acordos de Basileia, que corresponde ao montante em dívida na data de incumprimento da empresa/particular.

CCF – *Credit Conversion Factor* – parâmetro regulamentar estabelecido nos Acordos de Basileia, que corresponde à percentagem do montante não utilizado de uma linha de crédito com limite explícito, num dado momento de referência, que se prevê que seja utilizado até à data de incumprimento.

IRB – *Internal Ratings-Based approach* – designado por método das notação internas, foi estabelecido nos Acordos de Basileia e possibilita às IF a atribuição de notações internas (*ratings*) aos seus clientes e, consequentemente, a utilização de modelos de risco de crédito desenvolvidos internamente.

EBA – *European Banking Authority* – é a autoridade europeia responsável por assegurar a adoção das normas previstas nos regulamentos europeus, realizando uma supervisão prudencial de todo o setor bancário europeu. A EBA pertence ao Sistema Europeu de Supervisão Financeiras (SESF) que é composto por três autoridades de supervisão: Autoridade Europeia dos Valores Mobiliários e dos Mercados; a Autoridade Bancária Europeia (EBA) e Autoridade Europeia dos Seguros e Pensões Complementares de Reforma (EIOPA).

BIS – *Bank for International Settlements*.

*SME as retail* – definição estabelecida nos regulamentos europeus que possibilita às IF a classificação na carteira de retalho de empresas de pequena e média dimensão que detenham um total de créditos na IF inferior a 1 milhão de euros.



INE – Instituto Nacional de Estatística é o organismo oficial responsável por produzir e divulgar informação estatística sobre a realidade portuguesa.

RW – *Risk Weighted assets* – definição regulamentar dos Acordos de Basileia que corresponde à ponderação dos ativos em balanço da IF pelo seu nível de risco.

S&P – *Standard & Poors* – é uma agência de notação financeira pertencente ao Grupo McGraw-Hill, que publica análises sobre bolsas de valores mundiais e presta serviços de avaliação do risco de empresas, governos ou países em risco de incumprimento. As agências S&P, *Moody's* e *Fitch Ratings* são atualmente as principais agências de notação de risco.

LEQ – *Loan Equivalent Factor* – definição dada nos EUA ao fator de conversão, nomeadamente, nas normas regulamentares de implementação dos Acordos de Basileia.

FTH – *Fixed Time Horizon* – método de recolha de observações baseado num horizonte fixo, isto é, que observa o acréscimo na exposição com base na 12.<sup>a</sup> data anterior à data de incumprimento.

VTH – *Variable Time Horizon* – método de recolha de observações baseado num horizonte variável, isto é, que observa o acréscimo na exposição com base nos doze meses antecedentes à data de incumprimento.

$L_t$  – símbolo utilizado para descrever o limite contratado pela empresa à data de referência  $t$  de uma linha de crédito com limite explícito.  
 $E_t$  – símbolo utilizado para descrever o montante utilizado à data de referência  $t$  de uma linha de crédito com limite explícito. Este montante pode ser também designado por exposição, isto é, o montante em dívida a que a IF está exposta e que se encontra registado no balanço.

$EAD_i$  – símbolo utilizado para descrever o montante utilizado ou exposição na data de incumprimento de uma empresa  $i$ .

AUF – *Additional Utilization Factor* – medida do EAD utilizada por Valvonis (2008) que corresponde ao quociente da variação no montante utilizado entre a data de referência e a data de incumprimento e o limite da linha na data de referência.

RMSE – *Root Mean Square Error* – medida do desvio dos valores estimados contra os observados, dado pela raiz quadrada do desvio quadrático médio.

$H_0$  – hipótese nula.

$H_1$  – hipótese alternativa.

$Y_i$  – valor observado de uma variável dependente.

$CCF_i$  – valor observado da variável dependente CCF.

$\bar{Y}$  – média de uma variável dependente.

$\overline{CCF}$  – média da variável dependente CCF.

$\sigma_Y$  – desvio-padrão da variável  $Y$ .

$\rho$  – coeficiente de correlação de *Spearman*.

$\alpha$  – nível de significância – Erro Tipo I.

$B_j$  – parâmetros desconhecidos.

$\varepsilon$  – erro aleatório.

$\sigma^2$  – variância dos erros ou resíduos.

$COV(\varepsilon_i, \varepsilon_j)$  – covariância dos resíduos.

$SQ_{regressão}$  – soma dos quadrados dos erros devidos à regressão.

$SQ_{total}$  – soma quadrática total.

$SQ_{resíduos}$  – soma do quadrado dos resíduos.

$\hat{B}_j$  – estimadores dos parâmetros  $B_j$ .

$FIV$  – Fator de inflação da variância.

$R^2$  – Coeficiente de Correlação.

## ÍNDICE DE FIGURAS

Figura 1 – Perdas esperadas ( <i>Expected Loss</i> ) e perdas não esperadas ( <i>Unexpected Loss</i> )	9
Figura 2 – Comportamento da utilização de uma linha de crédito de um contrato que incumpriu	21
Figura 3 – Exemplo com 2 contratos (DF1 e DF2) do método de recolha de CCF observados (designados pelo autor «LEQ») a partir do Horizonte Temporal Fixo	22
Figura 4 – Exemplo para 1 contrato (DF1) do método de recolha de CCF observados (designados pelo autor por «LEQ») a partir do Horizonte Temporal Variável	23
Figura 5 – Exemplificação de um modelo de regressão linear múltipla (esquerda) e de um modelo de árvore de regressão (direita).	27
Figura 6 – Exemplificação de um modelo de regressão linear múltipla com duas variáveis independentes.	31

---

## ÍNDICE DE TABELAS

Tabela 1 – Análise Univariada dos coeficientes de correlação de <i>Spearman</i> , ao nível de significância de 5%- Amostra VTH restringida ao nível de utilização inferior a 99%	46
Tabela 2 – Análise Univariada dos coeficientes de correlação de <i>Spearman</i> , ao nível de significância de 5% – Amostra FTH restringida ao nível de utilização inferior a 99%	49
Tabela 3 – Sumário da <i>Performance</i> dos modelos de Árvores de Regressão – Amostra VTH restringida ao nível de utilização inferior a 99%	54
Tabela 4 – Sumário da <i>Performance</i> dos modelos de Árvores de Regressão – Amostra FTH restringida ao nível de utilização inferior a 99%	56
Tabela 5 – Coeficientes dos Modelos de regressão e Sumário da <i>Performance</i> dos Modelos – Amostra VTH restringida ao nível de utilização inferior a 99%	58
Tabela 6 – Coeficientes dos Modelos de regressão e Sumário da <i>Performance</i> dos Modelos – Amostra VTH restringida ao nível de utilização inferior a 85%	59
Tabela 7 – Coeficientes dos Modelos de regressão e Sumário da <i>Performance</i> dos Modelos – Amostra FTH restringida ao nível de utilização inferior a 99%	60
Tabela 8 – Coeficientes dos Modelos de regressão e Sumário da <i>Performance</i> dos Modelos – Amostra FTH restringida ao nível de utilização inferior a 85%	61
Tabela 9 – Sumário da <i>Performance</i> dos Modelos do método de Horizonte Temporal Variável com base nas médias por classes/grupos – Resultados na Amostra de Validação	63
Tabela 10 – Sumário da <i>Performance</i> dos Modelos do método de Horizonte Temporal Fixo com base nas médias por classes/grupos – Resultados na Amostra de Validação	64
Tabela 11 – Sumário da <i>Performance</i> dos Modelos do método de Horizonte Temporal Variável com base na Regressão múltipla – Resultados na Amostra de Validação	65
Tabela 12 – Sumário da <i>Performance</i> dos Modelos do método de Horizonte Temporal Fixo com base na Regressão múltipla – Resultados na Amostra de Validação	66
Tabela 13 – Sumário da <i>Performance</i> da média simples nos dois métodos de horizonte temporal na Amostra de Desenvolvimento restringida ao nível de utilização inferior a 99%	66
Tabela 14 – Sumário da <i>Performance</i> da média simples nos dois métodos de horizonte temporal	67

## ÍNDICE DE ANEXOS

<i>Anexo I – Indicadores de empresas não financeiras na Zona Euro .....</i>	<i>74</i>
<i>Anexo II– Taxas de Juro médias de empresas não financeiras na Zona Euro .....</i>	<i>74</i>
<i>Anexo III – Evolução mensal do Indicador Coincidente da Atividade entre 2000 e 2014 .....</i>	<i>75</i>
<i>Anexo IV – Evolução média mensal do CCF forçado ao intervalo unitário na Amostra global (período de 2010 a 2014).....</i>	<i>75</i>
<i>Anexo V – Análise dos CCF médios por tipologia de forçagem do CCF, por Pontuação e Nível de utilização na Amostra VTH após filtragem ao nível de utilização inferior a 99%.....</i>	<i>75</i>
<i>Anexo VI – Análise dos CCF médios por tipologia de forçagem do CCF, por Pontuação e Nível de utilização na Amostra FTH após filtragem ao nível de utilização inferior a 99%.....</i>	<i>76</i>
<i>Anexo VII – Sumário estatístico da Amostra VTH restringida ao nível de utilização inferior a 99% .....</i>	<i>77</i>
<i>Anexo VIII – Sumário estatístico da Amostra FTH restringida ao nível de utilização inferior a 99%....</i>	<i>78</i>
<i>Anexo IX – Descrição das variáveis utilizadas no desenvolvimento dos modelos em ambos os métodos de recolha.....</i>	<i>78</i>
<i>Anexo X – Análise gráfica da correlação das 6 principais variáveis explicativas com o CCF forçado ao intervalo [0,1] – Amostra VTH restringida ao nível de utilização inferior a 99% .....</i>	<i>80</i>
<i>Anexo XI – Análise gráfica da correlação das 6 principais variáveis explicativas com o CCF forçado ao intervalo [0,1] - Amostra FTH restringida ao nível de utilização inferior a 99%.....</i>	<i>81</i>
<i>Anexo XII – Análise gráfica das médias do CCF forçado ao intervalo [-1,1] e do CCF forçado ao intervalo [0,1] das Variáveis discretas Pontuação, Dimensão da Empresa, Colaterais e Indicador de Atividade - Amostra de VTH restringida ao nível de utilização inferior a 99%.....</i>	<i>82</i>
<i>Anexo XIII – Análise gráfica das médias do CCF forçado ao intervalo [-1,1] e do CCF forçado ao intervalo [0,1] das Variáveis discretas Pontuação, Dimensão da Empresa, Colaterais e Indicador de Atividade - Amostra de FTH restringida ao nível de utilização inferior a 99%.....</i>	<i>84</i>
<i>Anexo XIV – Resultados dos Modelos de estimação baseados em médias por classes (árvores de decisão) com o CCF forçado ao intervalo unitário e a [-1,1] (“Árvore de decisão com Floor -100%”) – Amostra VTH restringida ao nível de utilização inferior a 99%. .....</i>	<i>86</i>
<i>Anexo XV – Resultados dos Modelos de estimação baseados em médias por classes (árvores de decisão) com o CCF forçado ao intervalo unitário e a [-1,1] («Árvore de decisão com Floor -100%») – Amostra FTH restringida ao nível de utilização inferior a 99%. .....</i>	<i>89</i>
<i>Anexo XVI – Análise gráfica dos desvios relativos mensais no EAD dos modelos de regressão – Amostra VTH restringida ao nível de utilização inferior a 99%. .....</i>	<i>91</i>
<i>Anexo XVII – Análise gráfica dos desvios relativos mensais no EAD dos modelos de regressão – Amostra VTH restringida ao nível de utilização inferior a 85%. .....</i>	<i>92</i>
<i>Anexo XVIII – Análise gráfica dos desvios relativos trimestrais no EAD dos modelos de regressão – Amostra FTH restringida ao nível de utilização inferior a 99%.....</i>	<i>92</i>
<i>Anexo XIX – Análise gráfica dos desvios relativos trimestrais no EAD dos modelos de regressão – Amostra FTH restringida ao nível de utilização inferior a 85%.....</i>	<i>93</i>
<i>Anexo XX – Análise de resíduos dos modelos de regressão da Amostra VTH restringida ao nível de utilização inferior a 99%.....</i>	<i>93</i>
<i>Anexo XXI – Análise de resíduos dos modelos de regressão da Amostra VTH restringida ao nível de utilização inferior a 85%.....</i>	<i>95</i>
<i>Anexo XXII – Análise de resíduos dos modelos de regressão da Amostra de FTH restringida ao nível de utilização inferior a 99%.....</i>	<i>96</i>
<i>Anexo XXIII – Análise de resíduos dos modelos de regressão da Amostra FTH restringida ao nível de utilização inferior a 85%.....</i>	<i>98</i>
<i>Anexo XXIV – Desvios relativos mensais no EAD dos modelos de Horizonte Temporal Variável – Amostra de Validação .....</i>	<i>100</i>
<i>Anexo XXV – Desvios relativos mensais no EAD dos modelos de Horizonte Temporal Fixo – Amostra de Validação .....</i>	<i>101</i>

## ÍNDICE

1. Introdução	2
2. Revisão de Literatura e Enquadramento Regulamentar	3
2.1. Enquadramento Regulamentar	3
2.2. Anotações Teóricas	6
2.2.1. Características do crédito a PME condicionantes do CCF	6
2.2.2. Alternativas de modelização do CCF propostas na Literatura	8
3. Metodologia de Investigação	20
3.1. Horizontes Temporais	21
3.2. Definição matemática do Fator de Conversão	24
3.3. Modelos Matemáticos	27
3.3.1. Estimação com base na média de classes/ grupos	27
3.3.2. Regressão Linear Múltipla	29
4. Caso prático de estimação do EAD/CCF por Horizonte Temporal	36
4.1. Análise Exploratória das amostras	36
4.1.1. Recolha das amostras e tratamento dos dados	36
4.1.2. Características amostrais por Horizonte Temporal	40
4.2. Principais Correlações por Horizonte Temporal	44
4.3. Resultados dos Modelos de Estimação	51
4.3.1. Método de estimação com base em média de classes/grupos	51
4.3.2. Método de estimação com Regressão Linear Múltipla	56
5. Validação dos modelos	62
5.1. Análise de <i>Performance</i> dos Modelos	63
5.2. Nível de Conservadorismo	67
6. Conclusões	69
7. Bibliografia	72
8. Anexos	74

## 1. INTRODUÇÃO

O objetivo do presente documento é o de analisar para o portefólio de PME, com base nas técnicas de estimação mais utilizadas, o impacto nas estimativas do parâmetro EAD/CCF da utilização de metodologias de recolha da amostra alternativas, mais especificamente, no que respeita ao horizonte temporal utilizado.

Os eixos da referida análise centrar-se-ão, ainda, nos seguintes tópicos:

- 1) Variáveis mais relevantes para a estimativa do parâmetro EAD/CCF, no portefólio de PME por horizonte temporal;
- 2) Utilização de métodos estatísticos alternativos;
- 3) Inferência sobre os modelos mais adequados para a estimação do EAD/CCF no contexto regulamentar.

Com este objetivo, far-se-á inicialmente (Cap. 2.) uma revisão da literatura quanto aos estudos realizados mais relevantes sobre a estimação do EAD/CCF, bem como um breve enquadramento regulamentar dos requisitos gerais específicos para o cálculo deste parâmetro (EAD/CCF).

Em seguida (Cap. 3), são descritos os principais aspetos metodológicos que refletem a estrutura base do desenvolvimento dos modelos a comparar, nomeadamente o horizonte temporal para recolha da amostra, a definição do fator de conversão e as técnicas de estimação a utilizar para a construção dos modelos.

Com base numa amostra do portefólio de PME, apresentam-se posteriormente (Cap. 4) os resultados da estimação do parâmetro EAD/CCF, contemplando a análise descritiva da amostra por horizonte temporal, a apresentação das correlações com o CCF por horizonte temporal e os resultados da estimação com base nas árvores de decisão e na regressão linear múltipla, concluindo-se sobre a respetiva *performance* no atual contexto regulamentar.

Finalmente, procede-se à validação dos vários modelos desenvolvidos (Cap. 5), utilizando uma amostra *out-of-sample* e *out-of-time* que permita verificar a qualidade dos resultados ao nível da *performance* e do risco de subestimação do EAD.

## 2. REVISÃO DE LITERATURA E ENQUADRAMENTO REGULAMENTAR

### 2.1. ENQUADRAMENTO REGULAMENTAR

O quadro regulamentar europeu<sup>1</sup>, aplicável às instituições de crédito (IF) pertencentes aos estados-membros da UE, contempla a possibilidade de se utilizar o método das notações internas (*IRB*), permitindo às IF a estimação, com recurso a modelos internos, dos parâmetros de risco utilizados para efeitos de cálculo dos requisitos de capital regulamentar.

Para obter a aprovação por parte das autoridades competentes, para utilização de um modelo interno de risco de crédito sobre um parâmetro e para um determinado portefólio, a instituição deverá responder a todos os requisitos gerais e específicos estabelecidos neste quadro regulamentar, devendo refletir adequadamente o nível de risco desse portefólio.

No método IRB Avançado (IRBA) a IF deverá desenvolver modelos internos de risco e crédito para cada um dos seguintes parâmetros: (a) probabilidade de incumprimento (PD); (b) perda dado o incumprimento (LGD); (c) exposição no momento de incumprimento (EAD/CCF).

A probabilidade de incumprimento (PD) estima o risco de, no horizonte temporal de um ano, o devedor incumprir no pagamento do montante de crédito concedido.

A perda dado o incumprimento (LGD) estima o montante que não será recuperado caso o devedor incumpra. Este parâmetro representa a percentagem estimada da exposição no momento do incumprimento que não será recuperável.

O parâmetro EAD representa o montante estimado (i.e. exposição) que se considera que estará em risco no momento esperado de incumprimento. Nos contratos de natureza fixa, como os contratos de crédito de longo prazo, o EAD representa o montante atualmente em balanço, isto é, o montante do crédito vincendo acrescido dos juros periodificados. Nos contratos de natureza variável, o montante utilizado do crédito é registado em balanço e o montante não utilizado, correspondente à subtração do limite contratado pelo montante já utilizado, é registado fora de balanço (como extrapatrimonial). Tipicamente,

---

<sup>1</sup> Destaque para o Regulamento Europeu n.º575/2013 do Parlamento Europeu e do Conselho, de 26 de junho de 2013 (usualmente designado por «CRR – Capital Requirements Regulation»), que estabelece os principais requisitos prudenciais, no âmbito de capital regulamentar, aplicáveis às instituições de crédito, adoptando o quadro regulamentar de «Basileia III» neste domínio e sendo aplicável diretamente a todos os estados-membros da União Europeia.



quando um cliente está próximo de entrar em dificuldades financeiras, o montante utilizado tende a aumentar. Por esse motivo, o montante que estará em risco no momento do incumprimento deverá corresponder ao montante atualmente utilizado acrescido de uma percentagem do montante não utilizado que se espera vir a ser utilizado. Assim, para efeitos da modelização interna da estimação do parâmetro EAD/CCF, apenas os contratos de natureza variável são elegíveis.

Esta percentagem de acréscimo do montante utilizado, no enquadramento regulamentar europeu, designa-se por fator de conversão de crédito e representa<sup>2</sup> o «[...] rácio entre o montante atualmente não utilizado de uma linha de crédito que poderá ser utilizado em caso de incumprimento e ficará por conseguinte exposto a risco, e o montante atualmente não utilizado da linha de crédito [...]». Isto significa que a estimação do EAD é realizada por meio da estimação do parâmetro CCF. No método padrão (SA), este parâmetro não é estimado internamente e as IF que não receberem autorização para utilizar os seus modelos deverão aplicar os valores estabelecidos no regulamento europeu anteriormente referido. No caso específico das linhas de crédito com limites explícitos, o valor do fator de conversão a aplicar é de 50%, se o seu prazo de vencimento inicial for superior a um ano, e de 20%, se o prazo de vencimento for inferior<sup>3</sup>. Assim, modelos internos que, em média, apresentam estimativas inferiores permitem reduzir a necessidade de requisitos de capital.

Embora não esteja explicitamente definido na atual regulamentação, o conceito de CCF impõe implicitamente a necessidade de este parâmetro ser zero ou superior, considerando que deverá representar «o montante atualmente não utilizado de uma linha de crédito que poderá ser utilizado em caso de incumprimento». Esta interpretação não é, ainda assim, unânime entre estados-membros. No documento publicado pela EBA, em 17 de dezembro de 2013, *Report on the Comparability of Supervisory Rules and Practices*, o tratamento de CCF inferiores a zero é o principal fator de diferenciação entre as autoridades competentes<sup>4</sup>. Segundo o relatório, algumas autoridades optam por estabelecer um *Floor* de zero no CCF observado e outras nas estimativas finais por grupo/categoria, permitindo que o CCF possa ser negativo individualmente.

---

<sup>2</sup> V. Artigo 4.º alínea 56) do Regulamento Europeu n.º 575/2013.

<sup>3</sup> V. Artigo 166.º n.º10 alínea b) e c).

<sup>4</sup> Definição utilizada nos documentos regulamentares europeus que remetem para a autoridade que supervisiona a aplicação dos normativos. Em Portugal, a autoridade competente é o Banco de Portugal.

O artigo 182.º do regulamento europeu n.º 575/2013 estabelece os requisitos específicos de estimação do fator de conversão. Por sua vez, a EBA desenvolve as normas técnicas específicas para o cálculo dessas estimativas, sendo que as principais normas atualmente publicadas sobre o CCF encontram-se nos artigos 56.º, 57.º e 58.º do *Consultation Paper on Draft RTS on Assessment Methodology for IRB Approach* (EBA/CP/2014/36, de 12-nov-2014).

De acordo com as normas regulamentares atualmente em vigor, as IF deverão estimar os fatores de conversão com base nas médias observadas por grupos ou categorias, tendo em conta a média ponderada pelo número de casos<sup>5</sup> de incumprimento. A base de dados, para carteiras de retalho, deverá ser igual ou superior a 5 anos e deverão ser utilizados todos os incumprimentos registados<sup>6</sup>. As IF deverão ainda obter estimativas apropriadas para períodos de recessão económica (*downturn*), isto é, deverão ser utilizadas as médias deste período se forem mais conservadoras do que as médias de longo prazo. A natureza, a severidade e a duração da recessão económica deverão ser detalhadas pelas IF, embora se preveja a emissão para breve pela EBA de normas técnicas quanto a este tema. Por outro lado, quando se verificarem relações adversas entre os fatores económicos e a utilização das linhas, as IF deverão incorporá-las nas suas estimativas do fator de conversão.

A regulamentação atual não prevê qualquer norma técnica relativa à metodologia de recolha das amostras e à seleção dos fatores de risco, apenas alertando para a necessidade das escolhas estarem adaptadas ao tipo de exposição para o qual se aplicam. De acordo com o relatório da EBA *Report on the Comparability of Supervisory Rules and Practices*, as autoridades competentes também não põem obstáculos à utilização das várias metodologias existentes (método do *cohort*, método do horizonte temporal fixo, método do horizonte temporal variável, método *momentum*, etc.) ou na seleção dos fatores de risco (apenas uma autoridade competente estabeleceu como mínimo a segmentação por tipo de cliente e por tipo de produto).

As normas regulamentares permitem, assim, algum grau de liberdade no desenvolvimento dos modelos internos de estimação, embora alertem para a necessidade de prudência na

---

<sup>5</sup> Embora a definição do artigo 182.º não seja clara quanto à ponderação dos casos de incumprimento por número, o documento de consulta publicado pela EBA, *CP on Draft RTS on Assessment Methodology for IRB Approach*, explicita claramente no artigo 57.º alínea b) a obrigatoriedade de ponderar por número.

<sup>6</sup> V. artigo 57.º alínea (c) do *Consultation Paper* referido na nota anterior. Esta norma técnica parece remeter para a preferência do legislador em optar pela «forçagem» dos CCF negativos, em alternativa, à eliminação destes registos.

avaliação das mesmas, no sentido de se limitar «o impacto de uma recessão económica sobre os seus fundos próprios».

Recentemente, no sentido de diminuir a variabilidade dos RW entre as IF, o BIS publicou recentemente o documento consultivo *Reducing variation in credit risk-weighted assets – constraints on the use of internal model approaches* no qual propõe um conjunto de restrições à estimação do CCF, limitando por um lado o desenvolvimento de modelos internos a IF que sejam IRB *Foundation* e, por outro, impondo condições específicas na modelização do EAD/CCF às instituições que tenham recebido permissão para o efeito. Assim, no contexto das estimativas de EAD/CCF para posições de retalho, destacam-se algumas restrições propostas ao nível das práticas atualmente utilizadas nos desenvolvimentos dos modelos, nomeadamente: (a) a necessidade de que as IF garantam que tomaram medidas que permitem limitar os efeitos no EAD estimado de linhas com elevado nível de utilização; (b) a limitação da utilização de um *cap* no CCF observado, onde este pode assim ser superior a 100%; (c) a obrigatoriedade de utilizar o método de recolha da amostra baseado no horizonte temporal fixo. O mesmo documento propõe ainda a imposição de um *floor* nas estimativas de EAD, que deve corresponder à soma do montante atualmente utilizado (i.e. exposição registada no balanço) e 50% do CCF aplicável no método padrão sobre o montante não utilizado (i.e. exposição fora do balanço – *off balance sheet*).

## 2.2. ANOTAÇÕES TEÓRICAS

### 2.2.1. CARACTERÍSTICAS DO CRÉDITO A PME CONDICIONANTES DO CCF

#### *O Conceito de Pequena e Média Empresa (PME)*

A Comissão Europeia define o conceito de Pequena e Média Empresa no Artigo 2 da Recomendação n.º 2003/361/CE, de 20 de maio de 2003, o qual é aplicável às empresas com um número de empregados inferior a 250 e um volume de negócios anual que não exceda os 50 milhões de euros ou que apresente um balanço anual inferior a 43 milhões de euros. De acordo com esta definição, as PME representam na União Europeia 99,8%<sup>7</sup> do seu tecido empresarial, e onde aproximadamente 93% são microentidades. Em Portugal, de acordo com o boletim estatístico «Empresas em Portugal 2012», publicado

---

<sup>7</sup>Fonte: *Annual Report on European SMEs 2014/2015*, novembro de 2015, Comissão Europeia.

anualmente pelo INE, as pequenas e médias empresas<sup>8</sup> representavam 99,9% do tecido empresarial português em 2012.

Embora estas entidades incorporem a maioria do tecido empresarial, nem todas são elegíveis no conceito regulamentar de «*SME as Retail*». Este conceito, definido no Regulamento Europeu n.º 575/2013, permite que empresas PME possam ser consideradas na carteira de retalho para efeitos de cálculo dos requisitos de capital, desde que a sua exposição consolidada (na instituição financeira) seja inferior a 1 milhão de euros<sup>9</sup>. Recentemente, a União Europeia tem procurado criar condições que permitam reduzir as dificuldades de acesso ao crédito a que estas entidades estão sujeitas, especialmente em períodos de recessão da economia. Em 1 de janeiro de 2014 foi, por exemplo, introduzido um fator suporte que permite reduzir os requisitos de capital sobre exposições em PME, no sentido de incentivar as IF à concessão de crédito a estas entidades.

#### *Características deste segmento condicionantes do CCF*

As PME tendem a ser mais dependentes do setor bancário, pois não têm na maioria das vezes acesso ao mercado de capitais. De acordo com o relatório recentemente publicado pela EBA<sup>10</sup>, a principal fonte de financiamento utilizada pelas PME são as contas caucionadas<sup>11</sup>, representando 31% do total do financiamento nas microentidades, 38,9% nas pequenas entidades e 43,9% nas médias entidades. As linhas de crédito bancárias são, desta forma, a principal garantia de liquidez e de suporte à atividade operacional para estas entidades, sobretudo em períodos de reduzida liquidez.

Historicamente, as PME apresentam níveis de rendibilidade mais baixos comparativamente às grandes empresas<sup>12</sup>, sendo um dos indicadores de maior risco percecionados pelas IF, que adotam, assim, medidas restritivas na concessão e gestão do crédito.

De acordo com Sufi (2005), a flexibilidade de utilização das linhas gera problemas de agência mais severos, pelo que as IF tendem a conceder crédito a empresas com maior

---

<sup>8</sup>O conceito de PME adoptado pelo INE segue integralmente o conceito estabelecido na recomendação europeia.

<sup>9</sup> V. Artigo 147.º, n.º 2 e 5 do Regulamento Europeu n.º 575/2013 do Parlamento Europeu.

<sup>10</sup> V. *Discussion Paper* EBA/DP/2015/02, de 31 de julho de 2015

<sup>11</sup>As contas caucionadas são um tipo de linha de crédito com um limite explícito que se caracteriza por permitir montantes negativos na conta à ordem.

<sup>12</sup> Consultar, em Anexos, o Anexo I – Indicadores de empresas não financeiras na Zona Euro.

histórico de rendibilidade e a impor restrições contratuais de utilização das linhas associadas aos níveis de rendibilidade da empresa. Esta atuação tende a ser mais rigorosa quanto maior forem os custos de agência percebidos pela IF, ou seja, quanto menor for a informação disponível sobre a situação financeira da empresa. De facto, o mesmo estudo conclui que a violação de uma cláusula contratual está associada a uma redução de 30% na disponibilidade dos montantes não utilizados.

Assim, na concessão de linhas de crédito, as IF, adicionalmente à aplicação de taxas de juro e comissões mais elevadas<sup>13</sup>, impõem maiores restrições na flexibilidade de utilização dos montantes ainda disponíveis. A atuação das IF na redução dos limites ou no bloqueio do acesso aos montantes ainda disponíveis, provoca a redução da dívida no momento de incumprimento, sendo o principal fator do maior número de CCF negativos neste portefólio, comparativamente aos portefólios de Médias ou Grandes empresas. Este facto poderá assim implicar maiores dificuldades no desenvolvimento de modelos para estimativa daquele parâmetro, uma vez que, a distribuição de CCF será enviesada para a esquerda, isto é, com um elevado número de observações negativas.

#### 2.2.2. ALTERNATIVAS DE MODELIZAÇÃO DO CCF PROPOSTAS NA LITERATURA

Na perspetiva das empresas, as linhas de crédito com limites explícitos representam um empréstimo bancário flexível, orientado para o curto prazo, em que a IF permite à empresa a utilização de fundos, dentro de um limite previamente acordado. Neste sentido, as linhas de crédito são especialmente importantes diante das insuficiências temporárias de tesouraria ou na compensação de ruturas de liquidez inesperadas.

Na perspetiva das instituições financeiras, as linhas de crédito com limites explícitos apresentam vantagens perante outras tipologias de produtos. Em geral, permitem um retorno financeiro elevado, considerando que, além da taxa de juro sobre o montante utilizado, apresentam comissões sobre o montante não utilizado pelo devedor (registado como extrapatrimonial), permitindo ganhos adicionais diante de outras tipologias de contratos.

Aos ganhos gerados pelos juros e comissões são deduzidos os custos de financiamento e a perda esperada do risco assumido pela instituição na concessão de uma linha de crédito.

---

<sup>13</sup> Consultar, em Anexos, o Anexo II – Taxas de Juro médias a empresas não financeiras na Zona Euro.

Esta última é determinada pela multiplicação dos três parâmetros regulamentares: PD, LGD e EAD. No entanto, embora a perda prevista seja refletida nos resultados operacionais da IF, refletem uma média de perdas, que são inferiores às perdas originadas, por exemplo, em períodos económicos recessivos, como reflete a figura abaixo. Assim, dentro da regulamentação em vigor, às instituições é exigida a salvaguarda de uma parte do seu capital para absorção de perdas não esperadas. Estas perdas não esperadas são obtidas tendo em conta estes parâmetros, ou seja, pela multiplicação do EAD por uma função dos parâmetros PD e LGD.

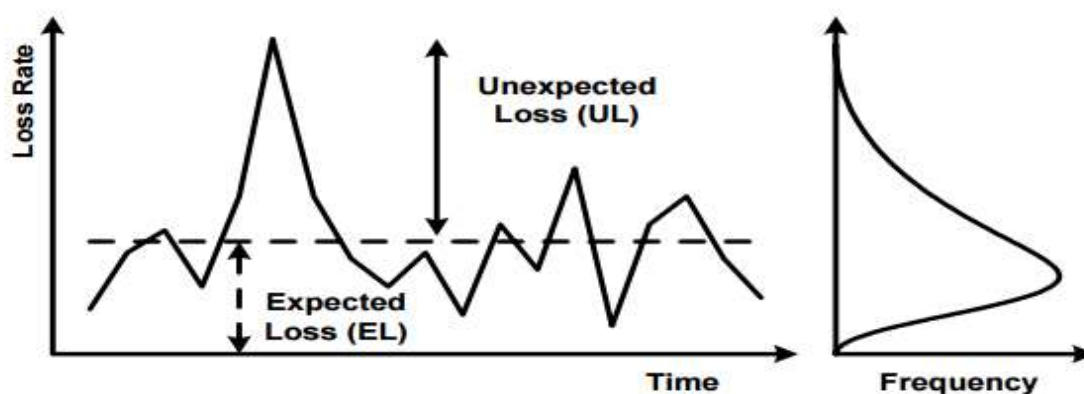


Figura 1 – Perdas esperadas (*Expected Loss*) e perdas não esperadas (*Unexpected Loss*); Fonte: *An Explanatory note on Basel II IRB Risk Weight Functions, Bank For International Settlements*

Desde 2004, o quadro regulamentar de Basileia II permite às instituições financeiras a modelização interna destes parâmetros. Posteriormente a este quadro, vários estudos surgiram sobre a estimação do parâmetro PD e outros, com menos frequência, sobre a estimação do parâmetro LGD. A estimação do EAD tem recebido, não obstante, menos atenção, especialmente no segmento de retalho, sobre o qual se desenvolve o presente trabalho. Contudo, ainda que o EAD tenha recebido menos enfoque, o fato de o EAD apresentar uma relação linear nos rácios de capital, ao contrário do parâmetro PD, implica que pequenas variações positivas tenham de ser compensadas na mesma proporção por Fundos Próprios. De facto, pela fórmula:

$$\text{Rácio de Capital} = \frac{\text{Fundos Próprios}}{RW * EAD}$$

$$\text{Onde } RW = (LGD * \left( N \left( \frac{1}{\sqrt{1-R}} * G(PD) + \frac{1}{\sqrt{1-R}} * G(0,999) \right) - LGD * PD \right) * 12,5 * 1,06^{14}$$

<sup>14</sup> V. Artigo 154.º, n.º1 do Regulamento Europeu 575/2013.

E:

- $N(x)$  designa a função distribuição de uma variável aleatória ( $x$ ) Normal Padronizada
- $G(Z)$  designa a inversa da função distribuição de uma variável aleatória Normal Padronizada
- $R$  designa o coeficiente de correlação calculado por:

$$R = 0,03 * \frac{1 - e^{-35*PD}}{1 - e^{-35}} + 0,16 * (1 - \frac{1 - e^{-3 * PD}}{1 - e^{-35}})$$

Uma variação de 5% no EAD implicaria um aumento de 5% no montante de Fundos Próprios para manter o mesmo nível de Rácios de Capital. Ainda assim, saliente-se que o limite máximo deste impacto depende, em grande medida, do nível de utilização das linhas de crédito. Isto é, portefólios com níveis de utilização das linhas muito elevados têm uma variação máxima no EAD mais baixa. A título de exemplo, um portefólio com um nível de utilização médio de 85% pode ter uma variação máxima de 17,6% no EAD (considerando um CCF conservador de 100%). Todavia, um portefólio com um nível de utilização de 65% terá uma variação máxima de 53,8% no EAD.

Outra razão para a menor atenção dada a este parâmetro é o facto de as bases de dados públicas disponíveis não conterem os fatores relacionais entre a instituição e a empresa (p.e. a identificação dos mecanismos de controlo das linhas descritos anteriormente), os quais são especialmente relevantes nas decisões a tomar aquando da modelização do parâmetro EAD/CCF.

Seguidamente, apresenta-se uma revisão sumária dos principais estudos realizados sobre a estimação do EAD/CCF e que foram consultados para a presente análise, enfatizando-se, sobretudo, as opções metodológicas tomadas no âmbito da recolha da amostra, tratamento de extremos, técnicas de estimação e variáveis utilizadas.

#### *Abordagem de Araten e Keisman (1994)*

O primeiro trabalho divulgado sobre a estimação do EAD em crédito a empresas utilizou dados do *Chase Manhattan Bank*, em 1994, e foi atualizado em 2003. O estudo foi conduzido com uma amostra de 104 contratos de crédito revogáveis, cujos *ratings* pioraram entre 1992 e 1993. Foi utilizado como medida de risco o CCF<sup>15</sup>. Os CCF foram estimados para contratos que incumpriram, sendo a amostra constituída apenas por

---

<sup>15</sup>O termo CCF utilizado neste relatório é equivalente ao termo LEQ referenciado nos artigos respeitantes às regras de implementação de Basileia II nos EUA. Por motivos de coerência e compreensão do presente relatório, optou-se por traduzi-lo para a expressão utilizada na UE, isto é, CCF.



*ratings* especulativos (iguais ou inferiores a BB). As principais conclusões indicaram que o CCF aumenta com a deterioração do risco de crédito do devedor.

*Abordagem de Asarnow e Marker (1995)*

Na sequência deste primeiro trabalho, Asarnow e Marker analisaram os padrões de utilização de linhas de crédito emitidas pelo *Citibank* de empresas com um histórico de *rating* atribuído pela S&P, com base em registos mensais durante o período de cinco anos, entre 1987 e 1991. As taxas de utilização foram calculadas com base na média de 84 meses para cada *rating*. Além da análise das taxas de utilização médias por categoria de *rating*, foram calculados os CCF, numa versão não publicada deste estudo, para 50 contratos onde os *ratings* deterioraram entre 1991 e 1993 (*ratings* especulativos, iguais ou inferiores a BB). Nos melhores *ratings*, os CCF foram extrapolados devido à falta de dados na deterioração do *rating* em *ratings* de investimento. Foram calculadas as médias trimestrais por *rating*. Nesta análise, ao contrário dos resultados obtidos no estudo do *Chase Manhattan Bank* (1994), o CCF apresentou uma relação inversa, diminuindo com o maior risco de crédito. O racional para esta relação pouco intuitiva poderá, em primeiro lugar, estar associada à fase de concessão do crédito, na medida em que devedores com menor risco estão sujeitos a menos cláusulas contratuais restritivas na utilização das linhas. Em segundo lugar, a monitorização do crédito pelo banco tende a ser menor em devedores com melhor *rating*, permitindo aos mesmos a utilização do remanescente das linhas sem a sua intervenção. Adicionalmente, outra explicação para este facto é o menor nível de utilização da linha nos melhores *ratings*, que proporciona maior variação entre o período de referência e a data de incumprimento.

*Abordagem de Araten e Jacobs (2001)*

Na sequência destes estudos, Araten e Jacobs analisaram 408 contratos de clientes do *Chase Bank* durante um período de seis anos, entre 1994 e 2000. As estimativas de CCF foram obtidas, a partir de um modelo de regressão para os contratos *revolving*<sup>16</sup>, tendo em consideração os fatores relacionados com as características do contrato, nomeadamente o *rating*, tempo (em anos) até ao incumprimento, localização, indústria, maturidade, tamanho da linha (limite contratado) e nível de utilização. A data de incumprimento correspondeu à data em que as empresas tinham recebido o pior *rating*, isto é, embora possam não ter incumprido tecnicamente foram sujeitas a reestruturação. As empresas

---

<sup>16</sup>De acordo com o artigo, foram incluídos contratos de crédito revogáveis de curto, médio e longo prazo.



que detinham mais de um contrato foram agrupadas num único contrato, para evitar redundância na base de dados. Os CCF observados correspondiam a observações trimestrais, no caso de se ter registado uma alteração no *rating*, e a anuais em caso contrário. Os CCF foram calculados até seis anos antes da data de incumprimento. Foram eliminados os registos com nível de utilização de 100%. No que respeita ao tratamento de valores extremos negativos, em particular, nas situações onde a exposição na data de referência era inferior à exposição na data de incumprimento, o valor de CCF observado foi forçado a 0 (alternativamente ao CCF negativo). A amostra tinha cerca de 28% dos casos nesta situação, ou seja, onde a dívida havia diminuído até à data de incumprimento. Os extremos positivos, gerados por situações em que o valor utilizado era superior ao limite, representavam 14% da amostra. Estes CCF observados superiores a 100% foram, assim, forçados a 100%. A média dos CCF observados era de 43,4% e o desvio-padrão de 41,4%, sendo que a distribuição dos CCF apresentava dois picos no 0% e 100%. Os resultados da regressão apenas incluíram as variáveis *rating* e o tempo até ao incumprimento. Ainda assim, a variável nível de utilização demonstrou boa capacidade discriminativa, embora muito correlacionada com a variável *rating*. A análise demonstra que a média do CCF vai aumentando com maior distância à data de incumprimento e com os melhores *ratings*. As restantes variáveis foram testadas, embora com resultados pouco significativos, o que se poderá dever à reduzida dimensão da amostra.

Este artigo é sobretudo importante na ênfase que atribui à necessidade de tratamento de dados previamente à obtenção das estimativas, detalhando todo o processo de exclusão e de tratamento dos valores extremos observados. Contudo, como uma das primeiras abordagens à estimação do CCF, os testes estatísticos estão pouco detalhados e as conclusões obtidas são, ainda assim, pouco robustas dado o reduzido número de observações, sobretudo nos melhores *ratings*.

#### *Abordagem de Moral (2006)*

O estudo de Moral é um dos primeiros estudos académicos realizados sobre a estimação do EAD, no contexto regulamentar, e aborda diversas problemáticas que surgem no desenvolvimento destes modelos.

O estudo relaciona as várias medidas de EAD que podem ser utilizadas apresentando as respetivas vantagens e limitações, bem como as metodologias de recolha da amostra, o

impacto do tratamento de valores extremos e a validação dos modelos no contexto regulamentar.

No que se refere às medidas de EAD, são tipicamente utilizadas duas medidas, o CCF e um fator sobre o limite. Note-se que embora o CCF seja obrigatório no atual quadro regulamentar europeu, outras medidas podem ser utilizadas no contexto de Basileia III. O autor sugere que embora o fator sobre o limite seja pouco utilizado, este apresenta algumas vantagens perante o CCF, nomeadamente pelo facto de, por um lado, estar definido quando o nível de utilização é de 100%, isto é, quando o limite é igual ao montante utilizado e, por outro, ser estável quando o nível de utilização está próximo de 100%. Estas vantagens poderão ser mais relevantes sobretudo em portefólios cujas linhas de crédito apresentam níveis de utilização muito elevados, considerando que uma percentagem significativa da amostra não tem CCF definido ou apresenta montantes não utilizados muito baixos que geram muitos valores extremos e, por isso, pouco informativos.

Na recolha da amostra de CCF observados, a metodologia utilizada apresenta, também, um impacto relevante das estimativas finais. Para obter estimativas de CCF aplicáveis a contratos que não estão em incumprimento, é recolhida uma amostra de CCF observados de contratos que incumpriram. A data de referência, na qual se observa o acréscimo na utilização da linha face ao momento de incumprimento, é assim relevante para as médias finais dos CCF observados.

Tipicamente, utilizam-se três metodologias distintas. O método de horizonte temporal fixo (FTH), o método do horizonte temporal variável (VTH) e o método *Cohort*. O método FTH observa o acréscimo de utilização exatamente T meses antes da data de incumprimento (geralmente utiliza-se T=12 meses antes). O método VTH observa o acréscimo ao longo dos T meses antes do incumprimento, isto é, no máximo poderá contemplar 12 observações de CCF por empresa. O método *Cohort* é um pouco mais complexo do que os métodos anteriores. A amostra é separada em vários *cohorts* e a observação do CCF representa o acréscimo entre o início de cada *cohort* e a data de incumprimento. Tanto os métodos FTH como *Cohort* não contemplam toda a informação disponível sobre a empresa, e a definição do período T, no primeiro método, e da extensão do *cohort*, no segundo método, é convencional. Em suma, o autor defende que as IF devem utilizar o método VTH, pois utiliza toda a informação disponível e assume que a

data de incumprimento poderá ocorrer a qualquer momento, embora exija às IF maior capacidade de guardar informação sobre os seus clientes.

Moral aborda ainda a problemática associada ao tratamento de valores extremos no CCF. O autor amplia a importância da limitação dos extremos negativos (originados pela diminuição da dívida), defendendo que o seu impacto é tanto maior quanto maior for o nível de utilização das linhas, ou seja, níveis de utilização elevados geram valores extremos muito grandes que não são compensados por extremos positivos (isto é, em que o CCF é superior a 100%). Assim, o autor analisa os diversos métodos de tratamento dos valores negativos, nomeadamente a censura de dados, forçagem das estimativas (i.e. forçando a que o valor estimado seja superior ou igual a 0%) ou a eliminação dos registos negativos. O autor conclui que a opção menos recomendável seria a eliminação, pois introduz inconsistências perante as amostras utilizadas para o cálculo do LGD, isto é, valores negativos de CCF muitas vezes estão associados a LGD elevados. No caso dos extremos positivos, o autor defende uma análise prudente às causas subjacentes. Assim, se o extremo é provocado pelo acréscimo de juros e comissões no EAD, esses extremos não deverão ser forçados a 100%, pois representam situações onde efetivamente a exposição final ultrapassou o limite contratado. Noutras situações, como alterações no limite não ocasionadas pela perceção de dificuldades financeiras pela IF, o autor considera relevante que na tomada de decisão seja salvaguardada a coerência destes casos com o cálculo das estimativas de LGD.

Quanto aos métodos estatísticos, o autor desenvolve a sua análise nos métodos mais utilizados pelas IF, nomeadamente as médias amostrais (com e sem ponderação) e os modelos de regressão simples sem constante, apresentando ainda como alternativa estimativas de CCF com base na regressão quantílica. O autor conclui que as médias simples devem conter como mínimo a variável explicativa «percentagem da linha ainda não utilizada», dada a elevada influência dos extremos negativos (gerados por linhas com montante utilizado próximo do limite) nas médias do CCF. Por outro lado, nos modelos de regressão simples sem constante, o autor utiliza as duas metodologias alternativas FTH e VTH e em ambas com dados originais e com censura dos CCF observados. Em ambas as metodologias de recolha da amostra, o autor conclui que é necessário analisar os valores extremos (tal como no método das médias simples), pois afetam, significativamente, os resultados da regressão mesmo no caso da censura de dados (isto é, um *Floor* de 0% e um *Cap* de 100%). Considera, ainda assim, que os resultados gerados

pela metodologia VTH são mais estáveis. Por último, o autor apresenta um modelo alternativo, cujo estimador é o quantil ótimo da distribuição de  $EAD_t - E_t$  condicionada aos valores de  $L_t - E_t$ . Segundo o autor, as estimativas geradas pela regressão quantílica são mais conservadoras do que as médias simples.

Adicionalmente, destaca-se que o autor defende a utilização de uma função perda que penalize as situações de subestimação do EAD, uma vez que têm maior impacto no contexto dos requisitos de capital. Segundo o autor, esta função permite às autoridades competentes avaliar o nível de conservadorismo implícito nas estimativas apresentadas pelas IF.

*Abordagem de Jimenez et al.(2006)*

O estudo de Jimenez utilizou dados da Central de Informação de Riscos de contratos de linhas de crédito superiores a 6000€ de todas as IF a operar em Espanha desde 1984. Numa primeira fase, foram analisadas as taxas de utilização com o objetivo de compreender o comportamento das empresas na utilização das linhas, em especial durante períodos de recessão económica. Numa segunda análise, foram desenvolvidos modelos de estimação do EAD. Esta análise permite retirar algumas conclusões importantes dada a extensão do período amostral, cerca de 20 anos, e a quantidade de informação disponível sobre as empresas.

No contexto da primeira análise, o autor conclui que empresas que incumpriram apresentam, em média, níveis de utilização das linhas mais elevados do que as empresas que nunca incumpriram. Esta diferença acentua-se à medida que o tempo até ao incumprimento das primeiras diminui, as quais atingem um nível de utilização médio de 90% no momento do incumprimento. Por outro lado, o autor sugere a existência de uma relação inversa entre o nível de utilização e o período económico, isto é, a utilização diminui em períodos de expansão e vice-versa.

Na segunda análise de estimação do EAD, foi utilizada a definição de incumprimento implementada à data, ou seja, 90 dias de atraso no pagamento ou a classificação pela IF de devedor com pagamento improvável. A amostra continha ainda informação detalhada da relação da IF com a empresa, por exemplo, duração da relação com a IF, o número de créditos totais, a percentagem de crédito total da empresa detida pela IF, etc. Para análise foram excluídas empresas que tinham mais de uma linha de crédito em diferentes IF. No

total, a amostra correspondia a 2 883 linhas de crédito de empresas que incumpriram e um total de 8 384 observações por ano e um nível de utilização médio de 51,5%.

O CCF médio foi calculado para um período até 5 anos antes do incumprimento, onde tal como em Araten e Jacobs (2001), o CCF diminui com o tempo até incumprimento, apresentando uma média final de 48,1%. Nesta amostra, o tamanho da linha é relevante e tem uma relação inversa, ou seja, maiores CCF estão associados a linhas com limites mais baixos. As linhas com menor maturidade, inferiores a um ano, e linhas sem colateral apresentam também maiores CCF. Note-se, no entanto, que na regulamentação em vigor as linhas de crédito com menor maturidade têm menor CCF regulamentar, que as linhas com maturidade superior a um ano.

O artigo é sobretudo importante nas evidências que apresenta quanto às diferenças observadas no nível de utilização de empresas incumpridoras e empresas que nunca incumpriram. No que respeita à estimação do EAD/CCF, é ainda assim pouco claro o tratamento dos extremos realizado (embora se possa presumir um tratamento idêntico ao estudo de Araten e Jacobs) e a metodologia de recolha da amostra (isto é, horizonte temporal utilizado).

#### *Abordagem de Valvonis (2008)*

O artigo de Valvonis, tal como em Moral, analisa vários aspetos a ter em consideração na estimação do EAD/CCF no contexto regulamentar. O autor parte das principais etapas para estimação do EAD/CCF desenvolvendo uma análise às várias opções. No tratamento de valores extremos, em alternativa aos estudos anteriores, Valvonis defende que no caso de extremos positivos originados por alterações no limite, apenas devem ser consideradas as observações de CCF após essa alteração. No caso dos extremos negativos, o autor defende que na geração de estimativas de CCF com base em médias simples, isto é, com base no seguinte método:

$$\widehat{CCF}_{Classe\ X} = MAX \left( 0; \frac{\sum_{i=1}^n CCF_i}{n} \right)$$

Dever-se-á colocar um *Floor* nos extremos negativos observados de -100%, no sentido de garantir a equivalência com o intervalo de variação positiva, isto é, de 0% a 100%. Assim, quando se verifica a diminuição da dívida, essa diminuição poderá representar no máximo o valor do montante ainda não utilizado. Quando os valores de CCF observados são muito grandes devido a denominadores próximos de 0 (isto é, níveis de utilização

próximos de 100%), estes devem, de acordo com o autor, ser forçados a 0% ou em alternativa a -100%, caso se utilize o método referido anteriormente.

Adicionalmente, quanto às técnicas de estimação, o autor defende que os modelos de regressão linear poderão não ser os mais adequados, uma vez que pressupõem uma relação linear entre as variáveis. Segundo o mesmo, devem ser utilizados métodos que permitam tomar em consideração a concentração das observações de CCF em 0% e 100%, como, por exemplo, regressões logísticas.

Neste estudo, foi ainda utilizada uma pequena amostra de 44 observações de cartões de crédito atribuídos a pequenas empresas que incumpriram entre 2005 e 2006 com o intuito de comparar os dois métodos: horizonte temporal fixo e *cohort*. A média de CCF do primeiro método era de 65,5% e do segundo de 17,6%. O nível de precisão na estimação do EAD medido pelo rácio da diferença do EAD observado e estimado e o EAD observado revelou-se melhor no segundo método, estando mais próximo de zero, embora esta precisão possa representar também maior risco de subestimação do EAD.

#### *Abordagem de Jacobs (2009)*

O estudo de Jacobs introduz uma análise empírica da utilização de diferentes medidas do EAD, comparando o CCF, com um fator sobre o montante utilizado e um fator sobre o limite (já referido por Moral). Segue ainda os princípios apresentados por Moral (2006), introduzindo uma função perda, que penaliza a subestimação face à sobrestimação. A amostra do estudo constituía 544 linhas de crédito de 496 empresas que entraram em incumprimento entre 1985 e 2007 e que em algum momento obtiveram *ratings* atribuídos pela Moody's ou S&P. A amostra correspondia a uma observação por empresa, 12 meses antes da data de incumprimento ou da data de alteração do *rating*, ou seja, foi utilizado o método de horizonte temporal fixo. No que respeita ao tratamento de extremos do CCF observado, foram introduzidas várias alternativas na forçagem dos valores, nomeadamente a forçagem para 0% e 100% e a forçagem para os valores do 5.º percentil e do 95.º percentil. A média do CCF forçado ao intervalo unitário era de 42,2% com um desvio-padrão de 41,0%. Por sua vez, a média do CCF da segunda opção era de 16,8% com um desvio-padrão de 210,4%. A mediana era de 33% em ambas as opções de tratamento dos valores extremos. Nesta amostra, o nível de utilização era de 45,9%. A análise univariada, com base no coeficiente de correlação de *Spearman*, demonstra, tal como nos estudos empíricos anteriores, que o *rating* e o nível de utilização têm maior

poder explicativo, com -0,178<sup>17</sup> e -0,335 respetivamente (considerando o CCF forçado ao intervalo unitário). Ainda assim, as variáveis «Dimensão da empresa», «Percentagem de Dívida coberta por colaterais» e «Qualidade do Colateral» revelaram-se significativas nesta amostra, com uma correlação de 0,178<sup>18</sup>, 0,1735 e 0,1594 respetivamente. Foram também analisados os efeitos de períodos de recessão económica nas médias de CCF, nomeadamente entre 1990 e 1991 e 2001 e 2002. Embora as médias se revelassem ligeiramente superiores no primeiro período de recessão (de 32,5% no período de 1985 a 1990 e 37,8% no período de recessão), o mesmo não sucedeu no segundo período (de 46,2% em 1993 a 2000 e 37% no período de recessão), pelo que não se podem retirar conclusões objetivas quanto aos seus efeitos no CCF.

No desenvolvimento do modelo CCF foram incluídas cerca de 11 variáveis e adicionalmente transformada a variável *rating* em 5 variáveis *dummy*, tendo sido incluídas as variáveis: nível de utilização, limite da linha, montante não utilizado, tempo até ao incumprimento<sup>19</sup>, dimensão da empresa, nível de intangibilidade do ativo, liquidez, resultados da empresa, qualidade do colateral, e três variáveis relacionadas com o endividamento da empresa.

Nos testes de validação sobre as três medidas de EAD testadas, o fator sobre o Limite apresentou o melhor resultado e o fator sobre o montante utilizado apresentou o pior resultado.

#### *Abordagem de Barakova et al. (2013)*

O estudo mais recente sobre a estimação do EAD realizado por Barakova *et al.* utilizou dados do programa *Shared National Credit* de linhas de crédito de 1300 empresas privadas e 300 empresas cotadas que incumpriram durante o período de 1997 e 2009. A base de dados continha informação anual dos limites das linhas e dos montantes utilizados, bem como dos *ratings* atribuídos pelas instituições financeiras (normalizados com recurso a uma escala de *rating* regulamentar comum). A medida do EAD é obtida

<sup>17</sup>Os *ratings* foram agrupados da seguinte forma: 1=AAA-BBB, 2=BB, 3=B, 4=CC-CCC e 5=D.

<sup>18</sup>Esta variável corresponde ao logaritmo do Capital Próprio 1 e 2 anos antes da data de incumprimento.

<sup>19</sup> Note-se que esta variável tem subjacente o fato de ser conhecida a data de incumprimento.

Considerando que estes modelos são aplicados em portefólios de contratos que nunca incumpriram, não é possível obter diretamente os valores desta variável.



pela diferença no montante utilizado entre o ano da data de incumprimento e o ano anterior a essa data. Foram utilizadas várias medidas do EAD, nomeadamente o CCF, o fator sobre o montante utilizado, um fator sobre o limite da linha e um fator, designado por AUF (*Additional Utilization Factor*), que corresponde ao quociente da variação no montante utilizado entre a data de referência e a data de incumprimento e o limite da linha na data de referência. O argumento para utilização desta medida prende-se com as suas propriedades numéricas, pois está bem definido para toda a amostra (neste estudo, 20% da amostra não está definida para as duas primeiras medidas de EAD). Para a definição do período de *downturn* foi calculada a média de classificações<sup>20</sup> da amostra, sendo selecionado o período onde a média de classificações estava um desvio-padrão acima. Adicionalmente, foram calculados os períodos de contração, isto é, períodos onde se registou uma subida das classificações.

Nesta amostra, cerca de 50% das empresas reduziram a dívida e apenas 4% a aumentaram. Por outro lado, cerca de 22% dos CCF têm um nível de utilização superior a 90%. Estes últimos contratos foram excluídos para efeitos do modelo preditivo, tendo o autor assumido um CCF de 0% nos testes de avaliação da *performance*. No que se refere ao tratamento de extremos, o autor analisa três opções, a forçagem dos CCF observados no 1.º e 99.º percentil, a forçagem a 0% e ao 99.º percentil e a forçagem ao intervalo unitário (0% e 100%). Contudo, na análise univariada apenas foram utilizados o CCF forçado ao percentil 1.º e 99.º e o CCF com *Floor* de 0% e *Cap* no 99.º percentil. Na estimação foi utilizada a regressão quantílica sem a censura dos CCF observados e a regressão multivariada com censura de CCF observados, como comparativo dos resultados obtidos no primeiro método.

Na avaliação dos modelos, o autor utilizou diferentes testes que, por um lado, medissem a *performance* e, por outro, a capacidade do modelo em prever acréscimos adicionais positivos no momento do incumprimento, em linha com as regras de Basileia II.

O autor conclui que as variáveis *rating*, nível de utilização, tamanho da linha e as contrações económicas têm impacto relevante nas estimativas de EAD, não havendo, no

---

<sup>20</sup>Note-se que classificações mais elevadas representam maior risco do devedor, isto é, maior probabilidade de incumprimento. As classificações estavam ordenadas por: 1-*Pass*, 2-*Special Mention*, 3-*Substandard*, 4-*Doubtful* e 5-*Loss*.



entanto, diferenças significativas por tipo de medida utilizada (CCF ou AUF) e por tipologia de tratamento de extremos (com ou sem *Floor* de 0%).

Nos testes de *performance* realizados com uma amostra *out-of-sample* e *out-of-period* e nos modelos com *Floor* não se registaram diferenças significativas. No que se refere ao problema de subestimação, o modelo de regressão quantílica, num quantil acima da mediana (55.º quantil), permitiu obter bons resultados em dados não censurados. No entanto, o modelo de regressão simples permitiu também obter os mesmos resultados considerando um *Floor* de 0%.

Em suma, os vários estudos realizados permitem retirar algumas conclusões importantes relativamente às variáveis que melhor explicam a variação do CCF e, por outro, identificam pontos importantes a ter em consideração no desenvolvimento de modelos de estimação do parâmetro CCF. No que respeita ao tratamento de extremos, os vários estudos sugerem alguma concordância na limitação dos extremos negativos, uma vez que na sua maioria optam pela forçagem a 0% ou a um determinado percentil da amostra, em alternativa a outras opções (como a sua eliminação). No que respeita ao horizonte temporal, a maioria dos estudos empíricos realizados utilizou o método de horizonte temporal fixo, o que poderá estar relacionado com a facilidade de acesso a informação, embora, de acordo com Moral, este método garanta menos estabilidade nas estimativas. Relativamente aos métodos estatísticos, embora se apresentem algumas alternativas, nomeadamente a regressão quantílica, os resultados apresentados nos modelos de regressão multivariada apresentaram um bom desempenho comparativamente a estes métodos. Nos testes de validação, verifica-se, sobretudo nos artigos mais recentes, uma preocupação crescente com a necessidade de avaliar os resultados dos modelos quanto ao risco da subestimação do EAD.

### 3. METODOLOGIA DE INVESTIGAÇÃO

Para uma determinada carteira de contratos de linhas de crédito com limites explícitos que não estão em incumprimento, o EAD estimado é determinado de forma indireta, por meio da soma do montante utilizado com o valor que resulta da multiplicação do montante atualmente não utilizado pelo CCF. Por sua vez, o CCF é estimado a partir de uma amostra de contratos representativos desta carteira que estão em incumprimento, permitindo obter um conjunto de CCF observados. O CCF observado de cada contrato em incumprimento

é obtido pelo quociente da variação da exposição (entre a data de incumprimento e uma determinada data de referência) e o montante não utilizado nessa data de referência.

A partir da amostra de CCF observados, recolhida de acordo com o horizonte temporal escolhido, são estimados os CCF com base nas características da empresa. A *performance* do modelo é, assim, avaliada a partir do EAD estimado, calculado nos termos atrás descritos.

### 3.1. HORIZONTES TEMPORAIS

O horizonte temporal condiciona a amostra de CCF observados e consequentemente as estimativas de EAD. Em primeiro lugar, deverá definir-se um horizonte que permita analisar o comportamento dos devedores que incumpriram na utilização das suas linhas. Uma vez que a estimativa de CCF deve ser consistente com a probabilidade de incumprimento (isto é, com o parâmetro PD), as estimativas de EAD devem referir-se à exposição dentro do período temporal máximo de 1 ano, como representa a figura abaixo.

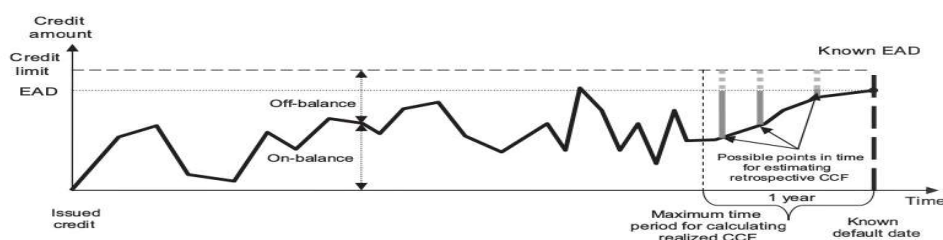


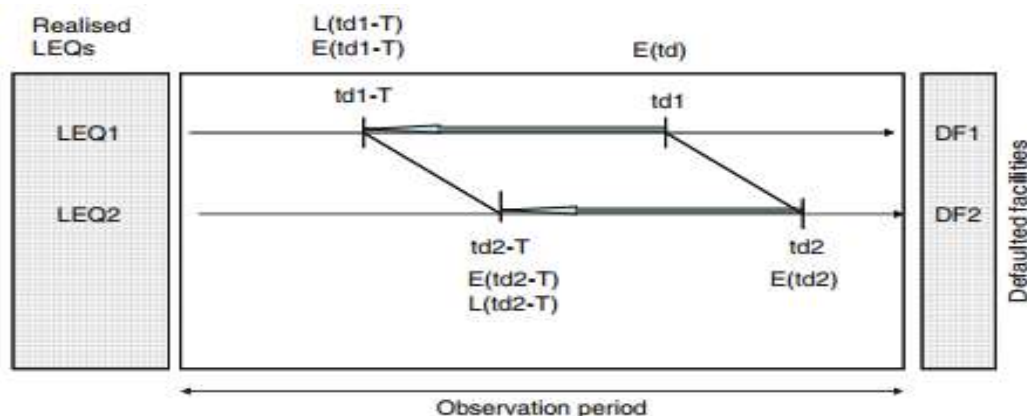
Figura 2 – Comportamento da utilização de uma linha de crédito de um contrato que incumpriu.<sup>21</sup>

Assim, a observação do acréscimo na exposição pode ser realizada em qualquer momento, mas dentro do horizonte temporal referido. No presente trabalho, serão utilizadas as duas principais metodologias com maiores diferenças entre si no que respeita os resultados das estimativas, nomeadamente: (a) o método de horizonte temporal fixo; (b) o método de horizonte temporal variável.

<sup>21</sup>Fonte: *Estimating EAD for retail exposures for Basel II purposes*, Vytautas Valvionis

### *Método de Horizonte Temporal Fixo*

No horizonte temporal fixo, em primeiro lugar é definido o horizonte T e posteriormente calculado, para cada contrato em incumprimento com nível de utilização  $\neq 100\%$ , o CCF observado pela fórmula:



$$CCF(td - T) = \frac{E(td) - E(td - T)}{L(td - T) - E(td - T)}$$

Figura 3 – Exemplo com 2 contratos (DF1 e DF2) do método de recolha de CCF observados (designados pelo autor por «LEQ») a partir do Horizonte Temporal Fixo<sup>22</sup>.

Este método permite obter uma boa dispersão das datas de referência e contribui para a homogeneidade dos CCF observados. A média de CCF é também superior, considerando um  $T=12$ , o que poderá contribuir para estimativas mais conservadoras. No entanto, não permite incluir contratos com antiguidade inferior a 12 meses e apenas contém uma observação por cliente, considerando implicitamente que a data de incumprimento irá ocorrer exatamente 12 meses após a data de cálculo da estimativa. Em carteiras com nível de utilização elevada, verifica-se uma grande incidência de valores extremos pouco informativos, o que em amostras pequenas poderá ter impacto na estabilidade das estimativas geradas.

### *Método de Horizonte Temporal Variável*

No horizonte temporal variável, após definido o horizonte T para cada contrato em incumprimento, é calculado um conjunto de CCF observados relativos a diferentes datas de referência anteriores (12 meses antes, 11 meses antes, etc.). As estimativas deste

<sup>22</sup> Fonte: *The Basel II Risk Parameters*, Capítulo XI - *EAD Estimates for Facilities with Explicit Limits*, Gregorio Moral.

método assumem que a data de incumprimento pode ocorrer em qualquer momento no horizonte T definido, tal como descrito na fórmula abaixo:

$$CCF(td - j) = \frac{E(td) - E(td - j)}{L(td - j) - E(td - j)}, \quad j = 1, 2, 3 \dots 12 \text{ meses antes}$$

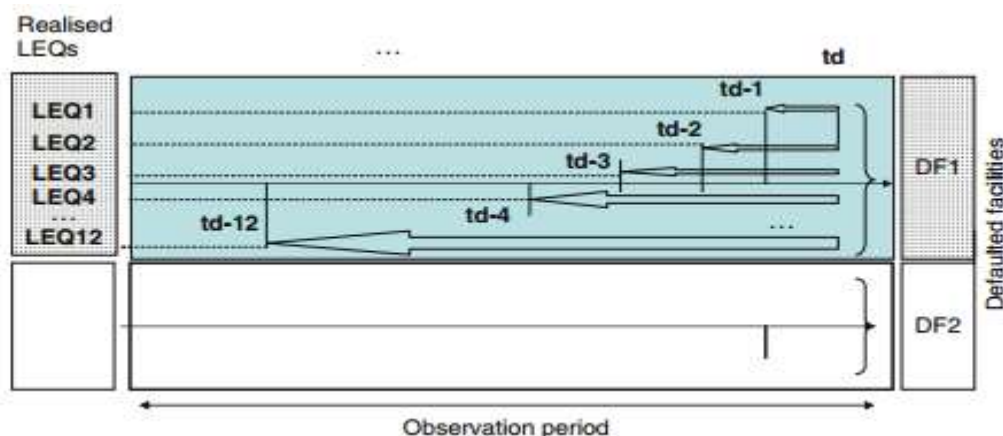


Figura 4 – Exemplo para 1 contrato (DF1) do método de recolha de CCF observados (designados pelo autor por «LEQ») a partir do Horizonte Temporal Variável<sup>23</sup>

Assim, a cada contrato estão associados, no máximo, 12 observações de CCF. Contudo, estas observações não são homogêneas, isto é, observações mais próximas da data de incumprimento apresentam CCF mais baixos comparativamente a observações de 12 meses antes. De acordo com Moral (2006), estas diferenças deverão ser reduzidas por meio de variáveis explicativas que permitam reduzir a variabilidade de CCF observados. Por outro lado, observações próximas do incumprimento devem ser analisadas, pois as IF tendem a restringir o acesso às linhas pouco antes do incumprimento, pelo que os CCF observados nestes casos não devem ser considerados na amostra.

Ao considerar várias observações por contrato, este método utiliza mais informação sobre o comportamento dos devedores, o que permite gerar modelos mais estáveis e com melhor capacidade de prever futuras utilizações adicionais.

Em suma, como referido anteriormente, não existe imposição regulamentar para a metodologia de recolha da amostra. No entanto, de acordo com os requisitos gerais devem ser incluídos «todos os dados, informações e métodos relevantes», sendo que «quanto menos dados a instituição tiver à sua disposição, tanto mais prudentes devem ser as suas

<sup>23</sup>Fonte: *The Basel II Risk Parameters, Capítulo XI - EAD Estimates for Facilities with Explicit Limits*, Gregorio Moral

estimativas»<sup>24</sup>. Assim, em portefólios onde apenas é possível obter amostras de reduzida dimensão, deverão ser utilizados métodos de recolha que garantam estimativas robustas e um maior grau de precisão, dado que em contrário, no contexto regulamentar, deverão ser ajustadas com uma «margem de prudência».

### 3.2. DEFINIÇÃO MATEMÁTICA DO FATOR DE CONVERSÃO

Nas linhas de crédito com limites explícitos, num determinado momento  $t$  o montante em exposição no momento de incumprimento estimado ( $\widehat{EAD}$ ) de um contrato  $g$  obtém-se pela soma do montante utilizado ( $E_t$ ) e o montante disponível ( $L_t - E_t$ ) multiplicado pelo fator de conversão estimado ( $\widehat{CCF}$ ), ou seja, o  $\widehat{EAD}(g)$  é expresso da seguinte forma:

$$\widehat{EAD}(g) = E_t + \widehat{CCF}(g) \times (L_t - E_t)$$

De acordo com a definição, o  $\widehat{CCF}$  representa uma estimativa da percentagem do montante não utilizado no momento  $t$  que estará utilizado no momento em que o devedor incumprir (momento  $td$ ). Para obter esta estimativa, as instituições utilizam amostras de contratos que incumpriram num determinado momento  $td$ , utilizando um horizonte temporal de observação ( $T$ ) que permite medir o acréscimo no montante utilizado entre o momento  $t$  (que pode ser o início do horizonte temporal definido) e a data de incumprimento  $td$ . No momento  $t$  a relação entre os fatores de risco do devedor ( $X_t$ ) (p.e. a pontuação, o tipo de contrato, os colaterais, as dívidas em atraso, etc.) e essa proporção do acréscimo observado ( $CCF_t$ ) permite obter as estimativas de CCF. Assim, uma vez definido  $T$ , os  $CCF$  observados da amostra são expressos como:

$$CCF_t = \frac{E_{td} - E_t}{L_t - E_t}$$

Onde o valor do numerador ( $E_{td} - E_t$ ) representa a variação, em montante, do montante utilizado entre a data de referência e a data de incumprimento, e o denominador o montante não utilizado ( $L_t - E_t$ ) na data de referência.

As estimativas de CCF representam, assim, o valor esperado da proporção do montante não utilizado que será utilizado em caso de incumprimento, condicionado ao horizonte temporal  $T$  e aos fatores de risco  $X_t$ , ou seja:

---

<sup>24</sup> V. Regulamento Europeu n.º 575/2013, artigo n.º 179, n.º1, alínea a) e alínea f).

$$\widehat{CCF}_{X_t} = E_t \left( \frac{E_{td} - E_t}{L_t - E_t} \mid td < T, X_t \right)$$

Pela sua definição, este fator apresenta algumas limitações na determinação das estimativas, nomeadamente:

- Não se encontra definido para linhas totalmente utilizadas  $L_t = E_t$ , o que implica que o EAD estimado destes contratos não pode ser obtido a partir das estimativas de contratos com nível de utilização igual a 100%;
- Não é estável quando o nível de utilização é muito elevado, tornando-se pouco informativo quando o montante disponível ( $L_t - E_t$ ) seja pequeno;
- Não considera alterações nos limites das linhas entre a data de referência e a data de incumprimento, o que muitas vezes é a causa de CCF observados superiores a 100%.

Adicionalmente, esta medida, como destacado por Moral (2006), é muito dependente do nível de utilização das linhas. Tomando a definição de CCF anterior e dividindo-a pelo limite ( $L_t$ ) obtemos:

$$CCF_t = \frac{\frac{E_{td}}{L_t} - \frac{E_t}{L_t}}{1 - \frac{E_t}{L_t}} = \frac{e_{td} - e_t}{1 - e_t}$$

Onde a derivada do CCF em ordem ao nível de utilização ( $e_t$ ) vem:

$$\frac{\partial CCF_t}{\partial e_t} = - \frac{1 - e_{td}}{(1 - e_t)^2}$$

Verifica-se que quanto menor o valor de  $(1 - e_t)^2$  maior será a variabilidade do CCF.

Tendo em conta as implicações do nível de utilização nos CCF observados, surgiram diversas abordagens no tratamento de valores extremos, destacando-se seguidamente as várias opções possíveis quando esses extremos são negativos (inferiores a 0%) ou positivos (superiores a 100%).

a) Tratamento de CCF observados muito inferiores a 0%

Tipicamente verifica-se uma elevada incidência de situações onde o  $E_{td}$  é inferior ao  $E_t$ , o que naturalmente provoca CCF observados muito negativos. Estes valores tendem a ser grandes extremos quando o nível de utilização é elevado (i.e. quando  $1 - e_t$  é pequeno).

No exemplo seguinte podemos ver o efeito no CCF observado do nível de utilização.

	$E_t$	$E_{td}$	$L_t$	$1 - e_t$	CCF
<b>Linha A</b>	100	80	110	9%	- 200%
<b>Linha B</b>	100	80	200	50%	- 20%

Considerando que o CCF estimado deverá, conceptualmente, ser igual ou superior a 0% (i.e.,  $E_{td} \geq E_t$ ) várias técnicas têm sido sugeridas no tratamento destes casos, com destaque para os seguintes:

1. Censura das observações com um *Floor* de 0%:

$$CCF_t = \frac{\max(0, E_{td} - E_t)}{L_t - E_t} \text{ ou } CCF_t = \frac{\max(E_{td}, E_t) - E_t}{L_t - E_t}$$

2. Censura das observações com um *Floor* baseado no percentil determinado pela análise da distribuição da amostra de CCF;
3. Eliminação das observações onde  $CCF_t < 0\%$ ;
4. Utilização da amostra não censurada e impondo uma restrição nas estimativas de CCF, ou seja, onde  $CCF_{(j)} \geq 0\%$ .

A primeira opção, embora represente um enviesamento da informação, é a mais utilizada pelas IF, uma vez que é simples de aplicar e permite introduzir algum grau de conservadorismo nas estimativas. A segunda opção, que consiste na imposição de um *Floor* baseado num percentil torna-o dependente do julgamento humano, contribuindo, assim, para a maior dificuldade de comparação dos resultados das estimativas entre as IF da União Europeia.

No presente trabalho, utilizou-se um *Floor* de 0% e um *Floor* de -100% (seguindo as sugestões de Valvonis, 2008). Uma vez que nesta amostra o elevado nível de utilização gerava uma elevada percentagem de extremos negativos, não se optou pelos restantes métodos, nomeadamente, um *Floor* baseado num percentil ou a eliminação de registos com CCF negativos.

- b) CCF observados superiores a 100%

Como referido por Moral (2006), os CCF superiores a 100% podem ser gerados por razões distintas que devem ser sujeitas a análise prévia. Introduzida a fórmula de cálculo anterior, verifica-se que se mantivermos constante o limite da linha e a exposição na data de referência inferior ao limite, o CCF torna-se superior a 100% quando a exposição no momento de incumprimento é superior ao Limite.

Assim, nas situações em que a exposição no momento de incumprimento inclui o montante de juros e comissões cobradas após o limite ter sido atingido, o CCF é mantido. Todavia, nos casos em que se verificou a diminuição do limite da linha, o CCF é forçado para 100%.

### 3.3. MODELOS MATEMÁTICOS

Como se referiu antes, o presente trabalho visa analisar o impacto dos métodos de recolha da amostra nas estimativas de EAD, a partir dos métodos estatísticos mais utilizados pelas IF, nomeadamente as médias por grupos ou classes, determinadas a partir de árvores de decisão e os modelos de regressão linear múltipla. Seguidamente, apresenta-se o desenvolvimento dos modelos matemáticos utilizados para a estimação do CCF.

Como referido anteriormente, o EAD é estimado a partir das estimativas de CCF, pelo que a *performance* dos modelos é avaliada no capítulo de «Validação dos Modelos». Na validação, utiliza-se uma amostra diferente da amostra de desenvolvimento e testam-se os resultados dos modelos sobre diferentes perspetivas.

#### 3.3.1. ESTIMAÇÃO COM BASE NA MÉDIA DE CLASSES/ GRUPOS

A estimação do CCF pode ser realizada por meio de árvores de decisão, tipicamente designadas por árvores de regressão, pois a variável resposta é contínua. Estas árvores são muitas vezes utilizadas pelas IF, pois não assumem uma relação linear entre as variáveis, não são sensíveis a extremos e são de fácil manutenção, interpretação e implementação nos processos internos de gestão de risco. No entanto, quando as relações entre as variáveis resposta e independente são lineares, a sua *performance* é facilmente ultrapassada por modelos de regressão linear, como representa a figura abaixo.

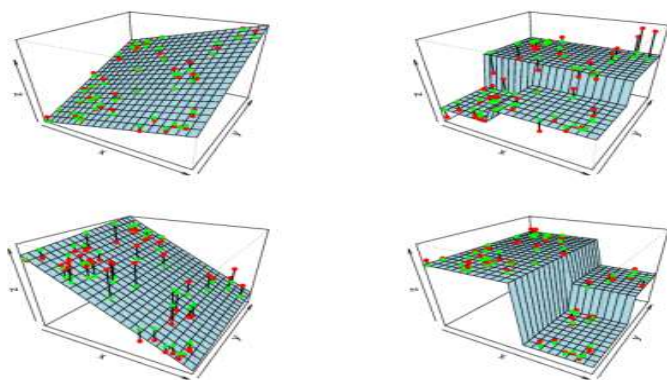


Figura 5 – Exemplificação de um modelo de regressão linear múltipla (esquerda) e de um modelo de árvore de regressão (direita).



Os vários algoritmos existentes para a construção de uma árvore de decisão dividem sucessivamente o conjunto de covariáveis  $X$ , criando grupos  $A_1, \dots, A_j$  com  $\hat{G} = j$  para todo o  $x \in A_j$ , de forma que sejam o mais homogêneos possível relativamente à variável resposta  $Y$ . Considerando que a variável resposta é contínua, isto é,  $CCF \in R$ , uma árvore de decisão divide o espaço  $X$  em regiões disjuntas  $A_k$  e gera um valor esperado  $E(CCF|X \in A_k)$  para cada região.

No desenvolvimento de árvores de decisão foi utilizada a ferramenta *SAS Enterprise Miner*. Em cada divisão a ferramenta calcula o teste F que avalia a utilidade dessa divisão quanto à variável resposta, ou seja, é selecionada a divisão que tem maior significância (i.e. o maior valor de  $-\log(\text{valor } p)$ ). O teste F representa assim o rácio da variabilidade entre os nós (calculado pela soma dos quadrados) e a variabilidade do nó. Isto é:

$$F = \left( \frac{SQ_{entre\ nós}}{SQ_{nó}} \right) \left( \frac{n-B}{B-1} \right) \sim F_{B-1, n-B},$$

onde:

$$SQ_{entre\ nós} = \sum_{i=1}^B n_i (\overline{CCF}_i - \overline{CCF})^2$$

$$SQ_{nó} = \sum_{i=1}^B SS_i = \sum_{i=1}^B \sum_{j=1}^{n_i} (CCF_{ij} - \overline{CCF}_i)^2,$$

e onde  $n$  é o número de observações e  $B$  o número de ramos.

Assim, o teste F permite identificar ramos cujos nós sejam o mais distintos possível e cujas observações dentro desses nós sejam o mais homogêneas possível. O teste é calculado recursivamente, de cada vez que é alterado o ponto de separação da variável de *input*. O tamanho da árvore é calculado automaticamente, isto é, o algoritmo para quando não é possível identificar uma variável de *input* cujas divisões resultem num valor  $p$  inferior a 1%, isto é, com um valor  $-\log(\text{valor } p)$  mínimo superior a 2.

Para a avaliação das árvores de decisão, obtidas a partir da amostra de desenvolvimento, foi calculada para cada árvore a média ponderada da soma dos quadrados dos erros entre o valor observado e o valor estimado de cada nó, dado pela fórmula:

$$\text{Erro Quadrático Médio} = \sum_{t=1}^T \frac{n_t}{n} \left[ \frac{1}{n_t} \sum_{i=1}^{n_t} (CCF_{it} - \overline{CCF}_t)^2 \right],$$

onde  $n_t$  é o número de observações do nó  $t$ ,  $CCF_{it}$  uma observação  $i$  da amostra de desenvolvimento incluída no nó  $t$  e  $\overline{CCF}_t$  o valor médio estimado do nó  $t$ .

Assim, são seleccionadas as árvores com o menor erro para o menor grau de complexidade possível.

### 3.3.2. REGRESSÃO LINEAR MÚLTIPLA

A regressão linear múltipla é a técnica mais utilizada para determinar a variação de uma variável dependente (ou resposta) face a variações de outras variáveis independentes. Estes modelos de regressão são muito utilizados pelas IF pois são fáceis de construir, simples de analisar e implementar nos seus sistemas de avaliação de risco.

#### *Análise Univariada – Coeficiente de Correlação*

Previamente à construção dos modelos de regressão foi analisada a correlação entre as variáveis dependente (CCF) e independente, no sentido de compreender o grau de associação entre as mesmas. Uma vez que o coeficiente de correlação de *Spearman* não pressupõe uma relação linear entre as variáveis, sendo também robusto a valores extremos, optou-se por utilizá-lo em desfavor do coeficiente de *Pearson*. Este coeficiente substitui os valores das variáveis por posições ordinais, determinando-se do seguinte modo:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n},$$

onde  $d_i$  é a diferença entre cada posição  $x_i$  e  $y_i$  e  $n$  o número de pares de observações.

#### *Introdução do modelo de Regressão Linear Múltipla*

Nos modelos de regressão múltipla a variável dependente  $Y$  é modelada como função linear de  $K$  variáveis independentes, sendo o modelo escrito como:

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + \varepsilon,$$

onde  $k$  corresponde ao número de variáveis explicativas e  $B_j$  os coeficientes de regressão. Para uma variável  $X_1$  a função representa uma linha reta, para duas variáveis  $X_1$  e  $X_2$  um plano e para  $K$  variáveis um hiperplano  $k$ -dimensional. Desta forma, a regressão múltipla pressupõe que a relação entre as variáveis independente e dependente é linear, pelo que a adição de novas variáveis independentes ao modelo deverá ter um efeito explicativo aditivo na variável dependente. Ainda assim, os modelos de regressão múltipla são bastante flexíveis, permitindo a introdução de variáveis transformadas, nos casos em que

a relação entre a variável resposta e a variável original não é linear. Este procedimento pode incluir não só transformações de uma, mas de várias variáveis independentes, incluindo, por exemplo, casos em que existe interação entre variáveis explicativas.

O ajustamento do modelo e a estimação dos coeficientes é feita com base num conjunto de observações da variável dependente e das correspondentes variáveis independentes, isto é:

$$Y_i = B_0 + B_1X_{i1} + B_2X_{i2} + \dots + B_kX_{ik} + \varepsilon_i,$$

com  $i = 1, 2, \dots, n$ , em que  $n$  representa a dimensão da amostra e os termos de erro  $\varepsilon_i$ 's variáveis tais que  $E(\varepsilon_i) = 0$ ,  $Var(\varepsilon_i) = \sigma^2$  e  $E(\varepsilon_i\varepsilon_j) = 0$ .

Nos modelos de regressão múltipla devem ser garantidos alguns pressupostos quanto aos resíduos ( $\varepsilon_i$ 's), nomeadamente: (a) o pressuposto de normalidade da distribuição; (b) média nula e (c) variância constante nas variáveis independentes incluídas no modelo e na variável resposta (i.e. a dispersão dos resíduos com a variável resposta e as variáveis independentes não devem apresentar qualquer padrão).

O modelo matemático com o qual se pretende obter as estimativas de CCF pode, assim, ser escrito como:

$$CCF_i = B_0 + B_1X_{i1} + B_2X_{i2} + \dots + B_kX_{ik} + \varepsilon_i,$$

para  $i = 1, 2, \dots, n$ . Este modelo também pode ser escrito em notação matricial considerando a matriz  $X$  das variáveis independentes, também designada matriz de planeamento, definida por:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{1k} \\ 1 & X_{21} & X_{22} & X_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & X_{n2} & X_{nk} \end{bmatrix}$$

Representando por  $CCF$  o vetor  $n \times 1$  cujo elemento  $i$  é  $CCF_i$ :

$$CCF = \begin{bmatrix} CCF_1 \\ \dots \\ CCF_n \end{bmatrix}$$

assim como  $B$  o vetor  $(k + 1) \times 1$  dos coeficientes,

$$\mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ \dots \\ B_k \end{bmatrix}$$

e  $\varepsilon$  o vetor  $n \times 1$  dos termos de erro:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_k \end{bmatrix}$$

Deste modo o modelo pode ser escrito como:

$$\mathbf{CCF} = \mathbf{XB} + \boldsymbol{\varepsilon}$$

Para obter as estimativas dos parâmetros  $B_0, B_1, B_2 \dots B_k$  é utilizado o Método dos Mínimos Quadrados.  $B_0$  representa o valor da constante e  $B_1, B_2 \dots B_k$  representam os valores dos coeficientes de regressão associados a cada variável independente. Os coeficientes de regressão  $B_k$  representam a variação esperada no CCF para cada unidade de variação em  $x_k$  quando todas as restantes variáveis se mantêm constantes.

Os estimadores obtidos por este método, também chamados estimadores de mínimos quadrados, são aqueles que minimizam a soma dos quadrados dos desvios verticais entre as observações e o hiperplano definido por estes mesmos coeficientes, como refletem os pontos vermelhos da figura 6.

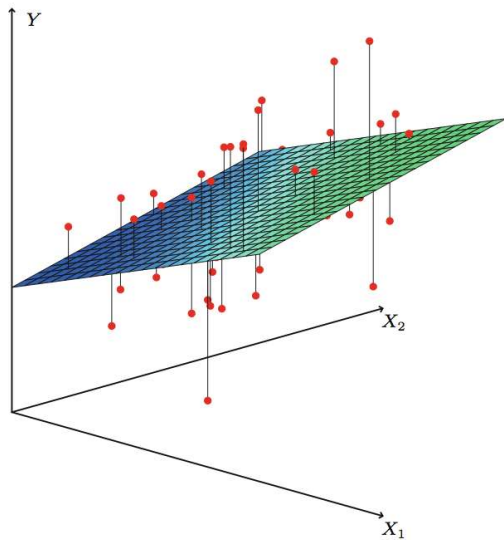


Figura 6 – Exemplificação de um modelo de regressão linear múltipla com duas variáveis independentes.

Deste modo, um modelo de regressão múltipla permite prever o valor do CCF em função de um conjunto de variáveis independentes,  $X^T = [1, X_1, X_2, \dots, X_k]$  e respetivos coeficientes de regressão  $B^T = [B_0, B_1, B_2, \dots, B_k]$  com o menor erro possível, isto é, minimizando a variância do termo de erro  $\varepsilon$ :

$$CCF = BX + \varepsilon$$

que para cada  $i$  é escrito como:

$$CCF_i = B^T X_i + \varepsilon_i,$$

em que  $X_i = [X_{1i}, X_{2i}, \dots, X_{ki}]$  representa o vetor das variáveis independentes relativas à observação  $i$ .

A partir desta fórmula, a estimativa dos coeficientes é obtida pela minimização de:

$$\sum_{i=1}^n (\varepsilon_i)^2 = \sum_{i=1}^n (CCF_i - B^T X_i)^2$$

O mínimo desta expressão pode ser obtido derivando-a em ordem a cada um dos coeficientes  $B_j$ ,  $j = 1, 2, \dots, k$  e igualando a zero. Assim, obtemos um sistema de  $k + 1$  equações a  $k + 1$  incógnitas dado por:

$$\sum_{i=1}^n (CCF_i - B^T X_i) X_{ij} = 0$$

De onde  $\hat{B}$ , o conjunto de estimadores para os coeficientes de regressão, vem como:

$$\hat{B} = (X^T X)^{-1} X^T CCF$$

Onde  $X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}$  e  $CCF = [CCF_1, CCF_2, \dots, CCF_n]^T$ .

Assim, o preditor ou valor ajustado de  $\widehat{CCF}_i$  é dado por:

$$\widehat{CCF}_i = \hat{B}_0 + \hat{B}_1 X_{i1} + \hat{B}_2 X_{i2} + \dots + \hat{B}_k X_{ik}$$

sendo a diferença entre a observação  $CCF_i$  e o respetivo valor ajustado (ou estimado)  $\widehat{CCF}_i$ , o resíduo ( $e_i$ ), isto é,  $e_i = CCF_i - \widehat{CCF}_i$ .

#### *Validação do modelo*

Após a estimação dos coeficientes, devemos verificar a qualidade do ajustamento do modelo aos dados, isto é, a variabilidade explicada por este.

Antes de verificar a qualidade do ajustamento, devemos ter em consideração os estimadores de  $\sigma^2$  e dos erros padrão. Como referido anteriormente, na regressão linear múltipla pressupõe-se que os resíduos  $\varepsilon_i$  com  $i = 1, 2, \dots, n$  são tais que:

- a)  $E(\varepsilon_i) = 0$ , onde  $E(\varepsilon) = 0$ ;
- b)  $V(\varepsilon_i) = \sigma^2$ , em que  $\varepsilon_i$  e  $\varepsilon_j$  não estão correlacionadas, isto é,  $COV(\varepsilon_i, \varepsilon_j) = 0$ ,  $i \neq j$ , onde a matriz de covariâncias do erro é dada por  $\sigma^2 I$ , em que  $I$  é a matriz identidade;
- c)  $\varepsilon_i \sim N(0, \sigma^2)$ .

Tendo em conta estes pressupostos, verifica-se que  $CCF \sim N(XB, \sigma^2 I)$  e que os estimadores de  $B$  não são enviesados, isto é,  $E(\hat{B}) = B$  e  $\Sigma(B) = \sigma^2 C$ , onde  $C = (X^T X)^{-1}$ . Deste modo para cada  $B_j$  temos que  $B_j \sim N(B_j, \sigma^2 C_{jj})$  em que  $c_{jj}$  representa o elemento  $j$  da diagonal da matriz  $C$ . Os estimadores das variâncias destes coeficientes de regressão são obtidos de um estimador de  $\sigma^2$ . O estimador  $\sigma^2$  é obtido a partir dos resíduos. Uma vez que a soma do quadrado dos resíduos é dada por:

$$SQ_{resíduos} = (CCF^T CCF - \hat{B}^T X^T CCF),$$

então prova-se que  $E(SQ_{resíduos}) = \sigma^2(n - p)$  e, portanto, um estimador não enviesado de  $\sigma^2$ .

Assim, a partir dos pressupostos de normalidade dos erros, em que  $\varepsilon_i \sim N(0, \sigma^2)$ , testa-se o nível de significância da relação linear entre as variáveis introduzidas no modelo e a variável resposta, CCF, isto é testa-se a hipótese:

$$H_0: B_1 = B_2 = \dots = B_k = 0$$

contra

$$H_1: B_j \neq 0 \text{ para algum } j, j = 1, 2, \dots, k$$

A rejeição da hipótese nula ( $H_0$ ) significa que pelo menos uma das variáveis incluídas no modelo contribui significativamente para explicar a variação da variável dependente (CCF). Usualmente, para este teste de hipóteses utiliza-se o teste F cuja estatística do teste é dada por:

$$F = \left( \frac{SQ_{regressão}}{SQ_{resíduos}} \right) \left( \frac{n-k-1}{k} \right) \sim F_{k, n-k-1},$$

onde  $n$  é o número de observações e  $k$  o número de variáveis independentes incluídas no modelo e em que a  $SQ_{regressão}$  se determina do seguinte modo:

$$SQ_{regressão} = \sum_{i=1}^n (\widehat{CCF}_i - \overline{CCF})^2$$

Note-se ainda que:

$$SQ_{total} = SQ_{regressão} + SQ_{resíduos},$$

em que:

$$SQ_{total} = \sum_{i=1}^n (CCF_i - \overline{CCF})^2.$$

Deste modo rejeitamos a hipótese nula se:

$$f_0 > f_{1-\alpha}[k, n - p]$$

com  $p = k + 1$  e  $\alpha$  o nível de significância. Isto é, rejeita-se  $H_0$  ao nível de significância  $\alpha$ , se o valor da estatística de teste  $F_0$  for superior ao valor do quantil de ordem  $1 - \alpha$  da distribuição  $F$  com  $k$  e  $n - p$  graus de liberdade.

Embora este teste permita verificar a significância do modelo de regressão, não indica se este é o mais adequado para estimar a variação do CCF em função das variáveis explicativas incluídas no modelo, além de que não permite retirar conclusões quanto às variáveis estatisticamente significativas para o modelo. Assim, para o efeito utilizam-se outros testes que permitam aferir a qualidade do ajustamento, comparando os seus resultados com outros modelos alternativos. No presente relatório foi utilizado o coeficiente de correlação que mede a proporção da variação do CCF que é explicada pela equação de regressão e é dado por:

$$C.Determinação = 1 - \frac{SQ_{Regressão}}{SQ_{Total}} = 1 - \frac{\sum(\widehat{CCF} - \overline{CCF})^2}{\sum(CCF - \overline{CCF})^2}$$

Por outro lado, de forma a avaliar a significância estatística de cada variável incluída no modelo é realizado um teste  $t$  sob a hipótese nula de que o coeficiente de regressão da variável é zero. A princípio a introdução de uma nova variável permite aumentar a soma dos quadrados da regressão e diminuir a soma de quadrados dos resíduos, no entanto, esse melhoramento poderá ser pouco relevante para o modelo, podendo até aumentar a média quadrática dos resíduos. A estatística de teste da hipótese nula é dada por:

$$T = \frac{\hat{B}_j}{se(\hat{B}_j)}, \text{ onde } se(\hat{B}_j) = \sqrt{\hat{\sigma}^2 C_{jj}},$$

sendo a região de rejeição dada por:

$$|t_0| > t_{1-\alpha/2}[n-p]$$

em que  $t_{1-\alpha/2}[n-p]$  representa o quantil de ordem  $1-\alpha/2$  da distribuição t de *Student* com  $n-p$  graus de liberdade. Quando não rejeitamos a hipótese nula, então o regressor  $X_j$  pode ser eliminado do modelo. Além deste teste, é realizado o teste F parcial, que avalia o impacto da inclusão da variável no modelo, sendo a estatística de teste determinada por:

$$F_{parcial} = \left( \frac{SQ_{inclusão\ da\ variável\ X_k}}{SQ_{resíduos}} \right) \left( \frac{n-k-1}{1} \right) \sim F_{1,n-k-1},$$

onde  $k$  é o número de variáveis independentes incluindo  $X_k$ . O teste F parcial pode ser generalizado para testar se vários coeficientes de regressão são nulos mas quando é utilizado apenas para um coeficiente, esse teste é equivalente ao teste t.

Na seleção das variáveis a incluir no modelo foi utilizado o método *Stepwise*. Neste método, primeiramente são adicionadas as variáveis cujo teste F parcial apresenta um valor  $p$  mais baixo. As variáveis vão sendo adicionadas se o resultado do teste F parcial tiver um valor  $p$  inferior a 5% (nível de significância). Após a adição de uma variável, as variáveis já incluídas são novamente avaliadas e removidas caso o valor  $p$  exceda o nível de significância definido.

Note-se que embora o pressuposto de normalidade dos resíduos seja fundamental para a estimação dos coeficientes, em grandes amostras é possível utilizar os testes de hipóteses anteriores sem que este pressuposto se verifique. De acordo com o teorema de Strivastava (1971) temos que se:

$$\max_{0 \leq i \leq 1} h_{ii} \rightarrow 0,$$

onde  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz  $X(X^T X)^{-1} X^T$ , também designada por *hat matrix*, então as estatísticas de teste sobre os coeficientes de regressão com distribuição  $F$  têm ainda, aproximadamente, esta distribuição.

Assim, de acordo com o teorema todos os testes anteriores podem ser realizados sem o pressuposto de normalidade dos resíduos desde que tenhamos  $h_{ii}$  pequenos. Embora seja difícil definir um número pequeno para os  $h_{ii}$  sobre o qual podemos, com segurança,



realizar estes testes (dado que depende não apenas do nosso julgamento crítico, mas também de como se apresenta a distribuição dos resíduos), regra geral assume-se que  $\max h_{ii} < 0,2$  é um bom indicador para a robustez dos resultados dos testes.

#### *Problema de multicolinearidade*

O problema de multicolinearidade verifica-se quando existem dependências fortes de 1 variável explicativa sobre outras 2 ou mais variáveis explicativas. Por outras palavras, a multicolinearidade verifica-se quando existe uma ou mais relações aproximadamente lineares em que cada uma dessas relações pode envolver 2 ou mais variáveis independentes. A existência de multicolinearidade tem efeitos negativos sobre as estimativas dos coeficientes de regressão e, por isso, na aplicabilidade do modelo, devem ser eliminadas as variáveis que estão na sua origem. Para detetar estas dependências foi realizada uma análise às variáveis com recurso ao método Fator de Inflação da Variância (FIV) calculado como:

$$FIV = \frac{1}{1-R_j^2},$$

onde  $R_j^2$  representa o coeficiente de determinação da regressão da variável explicativa  $j$  sobre as restantes variáveis explicativas.

De acordo com a fórmula, quando o valor de  $R_j^2$  é superior a 0.8 verifica-se o problema de multicolinearidade, pelo que a variável dependente foi eliminada.

#### 4. CASO PRÁTICO DE ESTIMAÇÃO DO EAD/CCF POR HORIZONTE TEMPORAL

##### 4.1. ANÁLISE EXPLORATÓRIA DAS AMOSTRAS

##### 4.1.1. RECOLHA DAS AMOSTRAS E TRATAMENTO DOS DADOS

Na recolha da amostra de contratos em incumprimento foi considerada a atual definição de incumprimento consagrada nos normativos europeus, especificamente no Regulamento Europeu n.º575/2013, bem como nas recomendações emitidas pela EBA. No âmbito destes normativos, considera-se em incumprimento uma empresa que mantém consecutivamente durante 90 dias uma dívida superior a 500 € ou sobre a qual se verifique um elevado grau de probabilidade de incumprimento das suas obrigações creditícias. Assim, recolhidas as empresas em situação de incumprimento foram selecionadas apenas as que detinham 1 ou mais contratos de linhas de crédito com limites explícitos. Estes

contratos referiam-se a linhas de crédito com um limite máximo de 1 milhão de euros, detidos por empresas que foram classificadas, na data de referência, como uma pequena e média empresa, de acordo com a recomendação europeia. A amostra contempla contratos que entraram em incumprimento entre o período de janeiro de 2010 a dezembro de 2014, perfazendo os 5 anos de histórico mínimo de dados estabelecido no regulamento europeu. A amostra de desenvolvimento contemplou os incumprimentos até dezembro de 2013 e a amostra de validação os incumprimentos ocorridos no ano de 2014.

Foram realizados alguns tratamentos adicionais aos dados. Nos casos em que a empresa detinha mais de 1 contrato, foram somados os limites e os montantes utilizados, para que a cada empresa correspondesse apenas uma observação de EAD, evitando a redundância da informação, conforme recomendado por Moral (2006). Adicionalmente, foram realizados alguns testes à qualidade de dados, tendo sido retirados os registos com limites nulos, EAD nulos ou contratos sem detalhe do tipo de empresa. A eliminação destes contratos visa garantir que as estimativas representam a realidade dos portefólios. No primeiro caso, tais situações referem-se a cortes totais das linhas que embora se mantenham nas bases de dados não caracterizam o comportamento deste portefólio, sendo que a sua inclusão iria representar CCF observados incoerentes (pois sendo o denominador sempre negativo um aumento da exposição teria um CCF negativo e uma diminuição da exposição um CCF positivo). O segundo caso por não representar a realidade do portefólio, onde se verificaria o pagamento da totalidade da dívida, foi também filtrado da amostra, pois a sua inclusão teria um impacto negativo nas médias do CCF (os quais seriam maioritariamente extremos negativos, que embora forçados a 0% iriam reduzir a média do CCF observado). No último caso, a eliminação deve-se à falta de informação quanto à dimensão da empresa, o que impossibilitava a garantia da definição «*SME as Retail*» adotada no presente relatório.

A amostra continha informação mensal sobre as características da empresa, as características do contrato, saldos de contas com a instituição e ainda informação financeira referente a 2 anos antes das datas de referência. Adicionalmente, foi também incluída informação mensal referente à situação da empresa face a dívidas e a pagamentos em atraso noutras instituições financeiras. Esta informação é disponibilizada pelo Banco de Portugal, na Central de Responsabilidade de Crédito.

Para a recolha das observações de CCF foi definido um horizonte temporal de observação de T=12 meses. Assim, no método de horizonte temporal fixo, considerou-se uma observação por empresa que correspondia ao 12.º mês anterior à data de incumprimento.

No método de horizonte temporal variável consideraram-se as 12 observações mensais anteriores à data de incumprimento. Note-se que a cada contrato não estão necessariamente associadas 12 observações, uma vez que em muitos casos verificou-se a existência de cortes na linha (i.e. limites nulos) meses antes do incumprimento. Estes casos, como referido anteriormente, foram excluídos da amostra. Finalmente, a amostra global, após os filtros aplicados, constituía 26763 observações, onde a média do nível de utilização era de 96,2% e a média do CCF forçado ao intervalo unitário e ponderado pelo número de observações era de 11,4%. A amostra de desenvolvimento, cujo período foi anteriormente referido, constituía aproximadamente 82% da amostra global.

Para definir o período de recessão económica (i.e. *downturn*), utilizou-se o Indicador Coincidente da Atividade desenvolvido pelo Banco de Portugal<sup>25</sup>. Este indicador varia com o PIB real e vários indicadores da atividade empresarial (p.e. o volume de vendas no comércio a retalho, índice de produção da indústria transformadora, etc.), permitindo obter uma perceção aproximada do estado da economia, sobretudo no contexto empresarial. Assim, para o período de recessão foram considerados os momentos onde este indicador era negativo. Pela análise deste indicador, desde 2000<sup>26</sup> verifica-se que o período de maior recessão coincide com a amostra de desenvolvimento do modelo, entre o início de 2011 e o fim de 2013, isto é, aproximadamente 83% do total das observações. Por outro lado, o EAD médio observado da amostra global aumentou apenas marginalmente no maior período de recessão, sendo que o CCF médio<sup>27</sup> também se manteve relativamente estável ao longo do período de 5 anos, aumentando ligeiramente apenas entre julho de 2011 e fevereiro de 2012 (com uma média de 13,4% face à média da amostra de 11,4%). Neste sentido, considerou-se que a atual amostra é representativa de um período de *downturn* acentuado, pelo que as estimativas obtidas são conservadoras.

No desenvolvimento dos modelos optou-se ainda por considerar apenas as linhas de crédito com nível de utilização inferior a 99%. A limitação da amostra a estas observações visou retirar os casos em que o limite era idêntico ao montante utilizado, os quais têm um CCF de 0% e os casos onde o montante utilizado era superior ao limite, que originavam CCF observados incoerentes (dado o denominador ser negativo). De facto, ao analisar a amostra de CCF, decompondo-o nas várias situações que podem ocorrer, verifica-se que

---

<sup>25</sup>V. *Um Novo Indicador Coincidente para a Economia Portuguesa*; Banco de Portugal; *Boletim Económico* - junho 2004, paginas 21 a 29.

<sup>26</sup>Consultar em Anexos, o Anexo III.

<sup>27</sup>Consultar em Anexos, o Anexo IV.

uma grande percentagem (cerca de 56% no horizonte temporal variável) correspondia a contratos com linhas totalmente utilizadas. Por outro lado apenas uma reduzida percentagem correspondia a aumentos efetivos na exposição (12,1% no horizonte temporal variável). Tal como referido por Moral (2006), verificaram-se mais casos de extremos negativos do que positivos, que têm um impacto significativo nas médias da amostra (-45.785,0% no VTH e -185.959,5% no FTH). No horizonte temporal variável, 18% das observações são reduções de dívida que provocam uma média de -288.190,5%, mas apenas 1,7% são situações onde o EAD supera o limite e onde a média é de 179 264,0%. Em suma, apenas 12,1% das observações representam efetivamente acréscimos na exposição onde a média se aproxima dos 71%. Do mesmo modo, no horizonte temporal fixo, apenas 15,3% acréscimos efetivos da exposição, embora a média seja mais elevada que no método anterior (aproximadamente 82%). Outro aspeto importante neste portefólio é o elevado nível de utilização médio em ambos os métodos, que poderá dever-se à elevada dependência das PME no financiamento bancário para suporte da sua atividade operacional. No horizonte temporal variável cerca de 83,7% das observações têm um nível de utilização superior a 95%, diminuindo ligeiramente no horizonte temporal fixo para os 78,6%. Em ambos os métodos, caso se optasse pela limitação da amostra ao nível inferior a 90% (tal como em Barakova *et al.*, 2013), ter-se-ia uma diminuição significativa de observações, o que poderia afetar a estabilidade das estimativas. Assim, embora se tenha testado o impacto de diferentes filtros nas médias do CCF observado, optou-se por utilizar o nível de utilização de 99% referido, pois garante a incorporação do maior número de observações, sobretudo no método de horizonte temporal fixo que contempla apenas uma observação por empresa. Não obstante, nos modelos de regressão testou-se o impacto nas estimativas da restrição da amostra aos contratos com nível de utilização inferior a 85%, dado que apresentava menor concentração de CCF negativos. Note-se que para efeitos de validação dos modelos assumiu-se um CCF de 0% onde o nível de utilização era superior a 99% ou, no caso dos modelos de regressão alternativos, a 85%. Contrariamente ao expectável, não se assumiu um CCF de 100% dado que esta opção iria reduzir o EAD estimado, pois nestes casos o limite é inferior ao montante utilizado.

#### 4.1.2. CARACTERÍSTICAS AMOSTRAIS POR HORIZONTE TEMPORAL

##### *Horizonte Temporal Variável*

A amostra de desenvolvimento final deste método de recolha continha 6223 observações após a restrição ao nível de utilização inferior a 99%. O CCF médio forçado ao intervalo unitário era de 25,7%<sup>28</sup> com um desvio-padrão de 0,5%. Por sua vez a média do CCF original era de -146,6% com um desvio-padrão de 6,8%. A média do CCF forçado ao intervalo  $[-1,1]$  era de -13,9% com um desvio-padrão de 1%. Embora se tenham eliminado muitos casos com CCF nulo, a mediana manteve-se nula, sendo que o CCF aumenta consideravelmente apenas no percentil 75, onde atinge os 53,7%.

O tratamento dos extremos é relevante se pretendermos desenvolver modelos com base na média. Assim, optou-se por 2 tipos de forçagem. O primeiro considerando a sugestão dada por Valvonis (2008), no qual a limitação a -100% poderia apresentar resultados satisfatórios colocando-se um mínimo de 0 nas estimativas de CCF. A segunda opção reflete a preferência das IF de limitar o CCF ao intervalo unitário. A primeira opção poderá não gerar bons resultados neste portefólio devido ao grande número de extremos negativos que influenciam consideravelmente as médias finais. Como referido anteriormente, estes extremos devem-se à redução da exposição. Ao analisar este comportamento por pontuação<sup>29</sup> verifica-se que as empresas com piores pontuações são as que mais diminuem a dívida até ao momento do incumprimento (cerca de 29% do total da amostra VTH). Como referido anteriormente, estes casos podem estar relacionados com a maior vigilância por parte das IF sobre os clientes com menor qualidade de crédito. Assim, perante a verificação de problemas de agência mais acentuados nas PME (i.e. onde existe pouca informação disponível sobre a situação financeira da empresa), a IF atribuiu-lhe piores *ratings*, exercendo maior limitação na disponibilidade das linhas. Na amostra VTH, cerca de 54% teve a pior pontuação e apenas 16% a melhor, o que explica o elevado número de casos com observações de CCF negativos. Por outro lado, apenas 2,9% constituem os casos mais severos, isto é, onde o EAD é superior ao Limite, o que sucede, em grande medida, devido à redução dos limites das linhas meses antes do incumprimento. Assim, dado que estes casos não correspondem efetivamente a situações onde crescem custos além do limite, assumiu-se uma forçagem a 100%. Note-se que não

---

<sup>28</sup> Consultar em Anexos, o Anexo VII.

<sup>29</sup> A variável *rating* continha 6 classes de risco que foram agrupados em 3 classes, designadas por pontuações. As classes foram reagrupadas no sentido de aumentar a homogeneidade dos CCF observados entre cada classe/pontuação. Assim a pontuação 1 corresponde aos melhores *ratings* e a pontuação 3 aos piores.

se optou pela forçagem a um intervalo baseado em percentis, por se considerar que a elevada concentração de grandes extremos negativos não permitia garantir médias conservadoras. Por exemplo, uma forçagem ao 5.º percentil teria um valor de -792,4% e ao 25.º percentil de -145%, ou seja, ainda abaixo dos -100% que geram ainda assim médias de CCF negativas.

O maior risco de acréscimos, tal como evidenciado em estudos anteriores, mantém-se nas linhas com níveis de utilização mais baixos e em empresas com melhor pontuação<sup>30</sup>. De facto, em níveis de utilização abaixo de 60% e na melhor pontuação as médias do CCF original e forçado estão bastante próximas, entre 63,9% e 63,3% (considerando o CCF forçado ao intervalo unitário), embora estes representem apenas 3% da amostra. Na melhor pontuação a aproximação entre o CCF forçado e o CCF original mantém-se até ao nível de utilização inferior a 85%, o que implica que nesta pontuação a média do CCF forçado ao intervalo unitário não está enviesada e representa o verdadeiro risco destas empresas. A partir deste nível de utilização, a distância entre as médias vai aumentando à medida que aumenta o nível de utilização, sendo que a média do CCF original torna-se muito negativa, de -325,1% no nível de utilização superior a 95%. Note-se ainda que a forçagem dos dados ao intervalo  $[-1,1]$  é também pouco conservadora nestes níveis de utilização elevados, pois as médias finais mantêm-se negativas. Nas pontuações mais baixas (pontuações 2 e 3), sobretudo na pior pontuação, o CCF original e forçado apenas se aproximam quando o nível de utilização é inferior a 60%, representando estes casos apenas 7,2% do total da amostra. Em níveis superiores a 85%, as médias do CCF tornam-se muito negativas nestas pontuações, representando conjuntamente 57,5% da amostra total. Embora pudéssemos, para efeitos de desenvolvimento dos modelos, desconsiderar estas observações assumindo um CCF conservador de 100%, optou-se por mantê-los garantindo maior estabilidade das estimativas.

O CCF médio vai também diminuindo à medida que a qualidade creditícia se deteriora, mesmo nos níveis de utilização mais baixos. Ao analisar o comportamento da utilização da linha, nomeadamente a média dos montantes utilizados e a média do EAD por nível de utilização e pontuação, verifica-se que na pior pontuação (pontuação 3) o EAD apenas supera o montante utilizado quando a utilização da linha está abaixo dos 60%, justificando assim a média do CCF original de 23,1%. Em níveis superiores o EAD mantém-se sempre

---

<sup>30</sup> Consultar em Anexos, o Anexo V.

próximo do montante utilizado, mas sempre inferior a este, pelo que um CCF médio forçado ao intervalo unitário de aproximadamente 17% garante algum grau de conservadorismo nestes casos. Por outro lado, na melhor pontuação, o EAD médio mantém-se superior ao montante utilizado até níveis de utilização de 85%, sendo o afastamento mais acentuado no nível inferior a 60%. Este facto está refletido nas médias do CCF original que se aproximam claramente do CCF forçado. A pontuação 2 é a mais estável, sendo o afastamento muito reduzido mesmo no nível de utilização inferior a 60%, onde o EAD é marginalmente superior ao montante utilizado, o que também se reflete nas médias do CCF original (45,3%).

#### *Horizonte Temporal Fixo*

A amostra de desenvolvimento final deste método continha apenas 682 observações, correspondentes ao 12.º mês anterior ao incumprimento e após a filtragem dos contratos com nível de utilização igual ou superior a 99%. Por ser um método que observa o acréscimo da exposição com 12 meses de distância à data de incumprimento, o numerador é superior, e por isso a média do CCF tende a ser também superior ao método de recolha de observações anterior. A média do CCF forçado ao intervalo unitário foi de 32,1%<sup>31</sup> com um desvio-padrão de 1,7%. No entanto, a média do CCF original foi inferior ao método variável, de -200,4% e o desvio-padrão superior de 22,2%. Da mesma forma, a média do CCF forçado ao intervalo [-1,1] foi inferior ao método VTH, de -15,8% com um desvio-padrão de 3,3%. A mediana do CCF original foi negativa, de -23,5%. Este método sugere maior dispersão nas observações de CCF, considerando que embora a mediana seja negativa o 75.º percentil tem um valor superior ao método variável, de 94%. Assim, embora as médias forçadas sejam mais conservadoras, a variabilidade é superior face ao método anterior.

A média do nível de utilização é apenas ligeiramente inferior, de 83% e o número de casos que reduziram a dívida superior, de 55%, o que explica a média de CCF original mais baixa face ao método variável. Sendo um método que não reflete atualizações de *rating* por parte da IF (considerando apenas um registo por empresa), existe uma maior percentagem de empresas com pontuação 2 (19% face a 13% no método variável) que reduz a dívida até ao momento de incumprimento. Assim, embora o padrão se mantenha, isto é, piores *ratings* representam a maioria das diminuições de dívida, esse padrão é mais

---

<sup>31</sup> Consultar em Anexos, o Anexo VIII.



claro no método variável, pois inclui atualizações de *rating* até à data de incumprimento. De facto, cerca de 42% teve a pior pontuação relativamente aos 54% da amostra anterior, o que se deve sobretudo à passagem da pontuação 2 para a pior pontuação durante os meses anteriores ao incumprimento. Verificou-se também um maior número de casos onde o EAD era superior ao limite, com 4,1% comparativamente aos 2,9% da amostra de horizonte variável.

Tal como no método variável, níveis de utilização abaixo de 60% e melhores pontuações apresentam as médias do CCF mais altas<sup>32</sup>, sendo também similares com e sem forçagem. As médias do CCF original apresentam-se superiores nas pontuações 2 e 3 face ao método variável, no entanto, são ligeiramente inferiores na pontuação mais alta (62,8% face a 63,9% no método anterior, considerando o CCF original). Contudo, à medida que aumenta o nível de utilização, nomeadamente a partir do nível de 90%, estas médias tornam-se bastante negativas e inferiores ao método variável, representando estes casos 58,7% da amostra total.

Tal como na amostra anterior, o EAD médio é superior ao montante utilizado médio apenas nos níveis de utilização mais baixos, sobretudo abaixo dos 60%. Acima destes níveis, o EAD médio encontra-se sempre muito próximo do montante utilizado e do limite.

Assim, considerando que as médias de CCF estão próximas nos 2 métodos, apresentando características de comportamento similares, poder-se-á considerar que o método de horizonte variável permite a inclusão de mais informação sobre as empresas, sobretudo no que respeita a empresas com tendência de maiores CCF, isto é, com menor nível de utilização e melhor pontuação. Adicionalmente, pelo fato deste portefólio apresentar uma elevada concentração de CCF negativos que são forçados a 0%, as amostras com poucas observações poderão apresentar estimativas menos robustas.

Por outro lado, a consideração do horizonte fixo permite, regra geral, obter estimativas mais conservadoras de CCF, melhor correspondendo ao interesse das entidades de supervisão bancária.

---

<sup>32</sup>Consultar em Anexos, o Anexo VI.



#### 4.2. PRINCIPAIS CORRELAÇÕES POR HORIZONTE TEMPORAL

A amostra continha um grande volume de variáveis, aproximadamente 100, pelo que se procedeu a um processo prévio de mineração de dados, seleccionando-se apenas as variáveis que demonstraram, através do teste de *Spearman*, maior correlação com a variável dependente. Assim, para cada tipologia de tratamento de valores extremos, descritos anteriormente, e para cada amostra, seleccionaram-se apenas as 14 variáveis com maior correlação<sup>33</sup>. No Anexo XIX apresentam-se as variáveis finais seleccionadas com maior correlação entre os vários CCF. Adicionalmente, foram ainda analisadas as diferenças nas médias do CCF por «Indicador de atividade», «Dimensão da empresa», «Pontuação» e «Colateralização». Na construção dos modelos de médias por classes/grupos a seleção das variáveis foi realizada iterativamente, introduzindo modificações nestas variáveis no sentido de melhorar os modelos. Seguidamente, sumarizam-se os resultados obtidos por método de recolha da amostra.

##### *Horizonte Temporal Variável*

Tal como demonstrado em estudos anteriores, de acordo com a Tabela 1<sup>34</sup> podemos verificar que o «Nível de utilização» é uma das variáveis com maior correlação com o CCF, sendo a correlação mais forte quanto maior o intervalo da distribuição do mesmo (coeficiente de -0,264 no CCF não forçado, de -0,243 no CCF forçado ao intervalo  $[-1,1]$  e de -0,210 no CCF forçado ao intervalo  $[0,1]$ ). Por sua vez, a variável «Montante não utilizado» apresenta uma correlação positiva com um comportamento entre os CCF similar à variável «Nível de utilização» (coeficiente de 0,214 no CCF não forçado, de 0,207 no CCF forçado ao intervalo  $[-1,1]$  e 0,180 no CCF forçado ao intervalo  $[0,1]$ ). A «Pontuação» é a variável com maior correlação no CCF forçado ao intervalo  $[0,1]$ , com um coeficiente negativo de -0,267. A informação interna da instituição sobre «Saldos de Depósitos da empresa em t» é também uma das variáveis mais importantes nos três CCF, embora superior no CCF forçado ao intervalo unitário (com coeficiente de 0,214). Sendo uma relação positiva, isto é, onde um saldo de depósitos mais elevado está associado a maiores CCF, podemos afirmar que também nesta amostra se verificam as conclusões de estudos anteriores de que os CCF tendem a ser superiores em empresas onde o risco de crédito é menos expectável pela IF. O «Rácio de depósitos totais/créditos totais» apresenta, tal como na variável anterior, uma correlação positiva e destacadamente

<sup>33</sup> Consultar em Anexos, a tabela com a descrição das variáveis utilizadas (Anexo IX)

<sup>34</sup> Consultar adicionalmente em Anexos o gráfico de correlações das principais variáveis (Anexo X)

superior no CCF forçado ao intervalo unitário (com um coeficiente de 0,186 que compara com um coeficiente de 0,134 no CCF forçado a  $[-1,1]$  e um coeficiente de 0,123 no CCF não forçado). Por outro lado, a variação verificada nos saldos de créditos totais que a empresa obteve na instituição apresentam também uma correlação positiva relevante com ambos os CCF. Assim, quanto maior o acréscimo no saldo de créditos maior tende a ser o CCF, sendo que a variação face a 12 meses antes, representada pela variável «Variação do saldo de créditos face a 12 Meses» apresenta a maior correlação em todos os CCF (de 0,134 no CCF não forçado, 0,143 no CCF forçado ao intervalo  $[-1,1]$  e 0,165 no CCF forçado ao intervalo  $[0,1]$ ). No mesmo sentido, a variável «Variação da dívida bancária total face a 6 meses antes» apresenta uma correlação positiva nos três CCF, sendo que em todos, esta é superior à variação face a 12 meses antes (de 0,121 no CCF não forçado, de 0,128 no CCF forçado ao intervalo  $[-1,1]$  e de 0,146 no CCF forçado ao intervalo  $[0,1]$ ). As variáveis relativas à informação financeira da empresa têm pouca expressão, o que pode estar relacionado com a distância da informação face à data de referência (igual ou superior a 2 anos). Apenas as variáveis «*Return on Equity*» e «Resultado Líquido em t-24» têm maior correlação nos três CCF, sendo que a correlação é em ambas as variáveis superior no CCF forçado ao intervalo unitário (de 0,116 e 0,106 respetivamente). No mesmo sentido que a variável «Saldo de depósitos da empresa em t», a relação positiva é contraintuitiva, pois um resultado líquido superior e um maior retorno no capital estão associados a maiores CCF. A variável «Resultados financeiros em t-24» apenas apresenta alguma correlação com o CCF não forçado (com 0,079) ou forçado ao intervalo  $[-1,1]$  (com 0,078).

CCF original		CCF forçado ao intervalo $[-1, 1]$		CCF forçado ao intervalo $[0, 1]$	
Variável	Coeficiente de Correlação de <i>Spearman</i>	Variável	Coeficiente de Correlação de <i>Spearman</i>	Variável	Coeficiente de Correlação de <i>Spearman</i>
Nível de Utilização	-0,26453	Pontuação	-0,26651	Nível de Utilização	-0,24273
Montante não utilizado	0,21371	Saldo de Depósitos da empresa em t	0,21384	Montante não utilizado	0,20678
Pontuação	-0,18344	Nível de Utilização	-0,20981	Pontuação	-0,19425
Saldo de Depósitos da empresa em t	0,15623	Rácio Depósitos Totais/Créditos totais	0,18584	Saldo de Depósitos da empresa em t	0,17051

Variação do Saldo de Créditos face a 12 meses antes	0,13387	Montante não utilizado	0,18036	Variação do Saldo de Créditos face a 12 meses antes	0,1426
Rácio Depósitos Totais/Créditos totais	0,12965	Variação do Saldo de Créditos face a 12 meses antes	0,16499	Variação do Saldo de Créditos face a 6 meses antes	0,13688
Variação do Saldo de Créditos face a 6 meses antes	0,12687	Variação do Saldo de Créditos face a 6 meses antes	0,15864	Rácio Depósitos Totais/Créditos totais	0,13659
Variação da dívida bancária total face a 6 meses antes	0,12144	Variação da dívida bancária total face a 6 meses antes	0,14615	Variação da dívida bancária total face a 6 meses antes	0,12838
Variação no Limite	0,1135	Variação da dívida bancária total face a 12 meses antes	0,13088	Variação no Limite	0,11159
Variação da dívida bancária total face a 12 meses antes	0,09851	<i>Return on Equity</i>	0,11635	Variação da dívida bancária total face a 12 meses antes	0,10669
Resultados Financeiros em t-24	0,07908	Resultado Líquido em t-2	0,10552	Resultados Financeiros em t-24	0,0778
Rácio Dívidas Bancárias em t-24/Passivo total em t-24	-0,07389	Variação no Montante Utilizado	0,10534	<i>Return on Equity</i>	0,07358
<i>Return on Equity</i>	0,06498	Saldo de Depósitos dos sócios em t	0,09708	Resultado Líquido em t-24	0,07216
Resultado Líquido em t-24	0,06189	Crédito Vencido à 1 Mês	-0,09389	Variação no Montante Utilizado	0,06985

Tabela 1 – Análise Univariada dos coeficientes de correlação de Spearman, ao nível de significância de 5%–Amostra VTH restringida ao nível de utilização inferior a 99%.

No que respeita às variáveis discretas apresentadas no Anexo XII, verifica-se que as médias tendem a ser mais díspares na variável ordinal, «Pontuação» e «Dimensão da empresa». As variáveis «Indicador de atividade» e «Colateralização» apresentam médias de CCF com diferenças pouco significativas entre categorias.

Ao nível da «Pontuação» verifica-se que as médias de CCF são sempre diferentes em todos os anos entre a melhor pontuação e a pontuação 2 e 3. Esta diferença agrava-se no ano de 2011, sendo este o ano onde o Índice Coincidente da Atividade apresentou também o valor mais baixo. Neste ano, a média do CCF forçado ao intervalo unitário atinge o máximo na melhor pontuação, 54%, sucedendo o mesmo no CCF forçado ao intervalo  $[-1,1]$  com um máximo de 30%. No entanto, na pontuação 2, o máximo é atingido no ano de 2010 em ambos os CCF. As médias da pior pontuação mantêm-se sempre estáveis ao longo dos anos, o que pode estar ligado ao controlo das linhas. Assim, nesta amostra, o efeito da crise económica poderá ter maior impacto nos CCF de empresas percecionadas

pela IF como de menor risco. Uma vez que nesta amostra (que se considera transmitir a realidade do portefólio de «SME as Retail» em Portugal), verifica-se um reduzido número de empresas com boa pontuação (16,5%), a média de CCF da amostra é pouco influenciada pelo seu comportamento.

No que se refere à «Dimensão da empresa», sendo este um portefólio de pequenas e médias entidades com linhas inferiores a 1 milhão de euros, as diferenças encontradas não são significativas. Ainda assim, as médias tendem a aumentar quanto maior a empresa, sendo as médias entidades as que apresentam maiores CCF, atingindo estas o máximo de 52% também no ano de 2011. Apenas em 2010, a média do CCF destas empresas é inferior às pequenas e microentidades, embora pouco se possa concluir devido ao reduzido número de observações de médias empresas neste ano. Assim, tal como evidenciado noutros estudos a «Dimensão da empresa» é relevante na diferença de médias do CCF, sendo que a média é superior no grupo das médias empresas comparativamente às pequenas e microentidades<sup>35</sup>.

A «Colateralização» das linhas tem também pouco influência no CCF. As médias são apenas ligeiramente superiores nas linhas com colaterais, atingindo o seu máximo no ano de 2013, com 35%. Nas linhas sem colaterais o máximo é atingido em 2011, com uma média de 27%. Este padrão mantém-se no CCF forçado ao intervalo  $[-1,1]$ , embora as médias sejam sempre negativas, com exceção dos contratos com colaterais no ano de 2013, cuja média foi de 6%.

No que respeita ao «Indicador de atividade», dado o reduzido número de observações de algumas categorias em determinados anos, optou-se por realizar a análise considerando a amostra total. Em ambos os CCF, o setor «Imobiliário» apresenta o maior risco de acréscimo, com 7,3% no CCF forçado a  $[-1,1]$  e 36,4% no CCF forçado ao intervalo unitário. As «Entidades Públicas» têm poucas observações, pelo que pouco se poderá concluir. Por outro lado, as atividades «Financeira», «Manufatura» e «Distribuição» apresentam as médias mais baixas no CCF forçado ao intervalo unitário (14,5%, 20,7% e 21,7% respetivamente), estando ligeiramente abaixo da média da amostra, de 25,7%.

---

<sup>35</sup> Note-se que foram ainda testadas as correlações das variáveis  $\log(\text{Ativo})$ , rácio de Ativo intangível sobre o Ativo e o valor do Capital Próprio. Contudo, estas apresentaram reduzida correlação, próxima dos 0.05, não tendo sido por isso incluídas nesta análise.

Em suma, nesta amostra baseada no método de recolha variável, verifica-se que o «Nível de utilização» e a «Pontuação» (ou *rating*) são as variáveis com maior relação com o CCF, embora com algumas variações no tratamento de extremos efetuado. Por outro lado, a informação interna que a IF dispõe quanto à situação de tesouraria da empresa revelou-se importante para o CCF, comparativamente com outras variáveis sugeridas em amostras de estudos idênticos. Contrariamente, a informação financeira disponibilizada pela empresa é pouco relevante, dada a distância da informação ao momento de referência. Quanto aos métodos de tratamento de extremos, o CCF censurado ao intervalo unitário sugere uma maior correlação com todas as variáveis, sendo exceção apenas o «Nível de utilização» (pelo facto de maiores níveis estarem associados a maiores extremos negativos que desta forma aumentam a correlação) e o «Montante não utilizado».

#### *Horizonte Temporal Fixo*

A amostra do método de horizonte temporal fixo embora contenha menos observações comparativamente ao método anterior apresenta correlações mais fortes. Contudo, não se verificam alterações significativas nas variáveis selecionadas, mantendo-se a respetiva ordem de escolha. De acordo com a Tabela 2, a variável «Nível de utilização» manteve-se a mais correlacionada com os três CCF, sendo que, tal como na amostra anterior, a correlação era tanto superior quanto maior o intervalo da distribuição do CCF (com um coeficiente de -0,431 no CCF não forçado, -0,384 no CCF forçado ao intervalo  $[-1,1]$  e -0,306 no CCF forçado ao intervalo  $[0,1]$ ). O «Montante não utilizado» apresenta também uma correlação significativa, mas sobretudo nos CCF com valores negativos, isto é, no CCF não forçado (0,364) e no CCF forçado ao intervalo  $[-1,1]$  (0,334). Neste método, o «Saldo de Depósitos da empresa em  $t$ » surge em primeiro lugar comparativamente à «Pontuação», com uma correlação similar entre os CCF, embora ligeiramente superior no CCF forçado ao intervalo unitário, com um coeficiente de 0,299. A «Pontuação» surge em terceira posição como a variável mais correlacionada com o CCF forçado ao intervalo unitário (com um coeficiente de -0,284), e em quarta quanto ao CCF não forçado (com 0,286) e o CCF forçado a  $[-1,1]$  (com 0,286). A «Variação do saldo de créditos face a 12 meses antes» é também uma das variáveis mais correlacionadas, sendo a correlação também superior no CCF forçado ao intervalo unitário (com um coeficiente de 0,236). As variáveis «Rácio Depósitos Totais/Créditos totais», «Variação da dívida bancária total

face a 6 meses antes» e «Variação do saldo de créditos face a 6 meses antes» encontram-se muito próximas entre si nos três CCF, sendo as mais correlacionadas, seguidamente às já referidas. Esta sequência é coerente com o método anterior, embora a correlação em todas as variáveis seja superior. As variáveis contendo a informação financeira da empresa surgem também no fim da lista. Este método diverge do anterior apenas ao selecionar as variáveis «Limite» e «Variação no montante utilizado». Esta relação é contudo fraca nos 3 CCF, sendo próxima de um coeficiente de 0.1.

CCF original		CCF forçado ao intervalo $[-1, 1]$		CCF forçado ao intervalo $[0, 1]$	
Variável	Coeficiente de Correlação de <i>Spearman</i>	Variável	Coeficiente de Correlação de <i>Spearman</i>	Variável	Coeficiente de Correlação de <i>Spearman</i>
Nível de Utilização	-0,43118	Nível de Utilização	-0,38365	Nível de Utilização	-0,30553
Montante não utilizado	0,36379	Montante não utilizado	0,33991	Saldo de Depósitos da empresa em t	0,29898
Saldo de Depósitos da empresa em t	0,28587	Saldo de Depósitos da empresa em t	0,28629	Pontuação	-0,28409
Pontuação	-0,24499	Pontuação	-0,24975	Montante não utilizado	0,27529
Variação do Saldo de Créditos face a 12 meses antes	0,20535	Variação do Saldo de Créditos face a 12 meses antes	0,20763	Variação do Saldo de Créditos face a 12 meses antes	0,23606
Rácio Depósitos Totais/Créditos totais	0,1992	Variação da dívida bancária total face a 6 meses antes	0,19374	Variação da dívida bancária total face a 6 meses antes	0,21822
Variação do Saldo de Créditos face a 6 Meses	0,18829	Rácio Depósitos Totais/Créditos totais	0,19027	Rácio Depósitos Totais/Créditos totais	0,20645
Variação da dívida bancária total face a 6 meses antes	0,17972	Variação do Saldo de Créditos face a 6 meses antes	0,18398	Variação do Saldo de Créditos face a 6 meses antes	0,20188
Variação da dívida bancária total face a 12 meses antes	0,13273	Variação da dívida bancária total face a 12 meses antes	0,14619	Variação da dívida bancária total face a 12 meses antes	0,16836
Saldo de Créditos da empresa em t	0,12859	Saldo de Créditos da empresa em t	0,13388	Saldo de Créditos da empresa em t	0,12407
<i>Return on Equity</i>	0,10363	Limite	0,11005	Variação no Montante Utilizado	0,12261
Variação no Montante utilizado	0,10341	<i>Return on Equity</i>	0,10937	<i>Return on Equity</i>	0,11311
Limite	0,10247	Variação no Montante utilizado	0,10574	Saldo de Depósitos dos sócios em t	0,09644
Resultado Líquido em t-24	0,08976	Caixa e depósitos bancários em t-24	0,09324	Limite	0,09281

Tabela 2 – Análise Univariada do coeficientes de correlação de *Spearman*, ao nível de significância de 5% – Amostra FTH restringida ao nível de utilização inferior a 99%.

Paralelamente à amostra anterior, ao analisar as variáveis discretas apresentadas no Anexo XIII, as principais conclusões mantêm-se, isto é, a variável «Pontuação» apresenta a maior diferença de médias, sendo que as restantes apresentam algumas diferenças, mas sobre as quais não é possível retirar conclusões objetivas.

A variável «Pontuação» é consistente ao longo dos anos em todas as pontuações, sendo que as médias decrescem à medida que diminui a pontuação. No entanto, na «pontuação 1» verifica-se que também no ano de 2011 a média do CCF foi superior, atingindo o máximo de 59% no CCF forçado ao intervalo unitário e 30% no CCF forçado a  $[-1,1]$ . Ainda assim, as médias são mais homogêneas entre anos comparativamente ao método de recolha variável. Esta relação mais homogênea pode explicar a maior correlação desta variável com o CCF face ao método anterior.

A variável «Dimensão da empresa», tal como no método anterior, apresenta diferenças pouco significativas ao longo dos anos, com exceção do ano de 2011, no qual se verifica uma diferença relevante entre as médias entidades e as pequenas e microentidades. Embora este facto suceda no método anterior, é mais acentuado nesta amostra. As médias entidades atingem em 2011 o máximo de 66% que compara com os 36% nas pequenas e 32% nas microentidades, considerando o CCF forçado ao intervalo unitário. No entanto, nos anos de 2010 e 2012 este padrão inverte-se, embora com menos diferenças entre os grupos. Nesta amostra, as observações de médias empresas representam apenas 4,5% do total, pelo que o efeito desta variação na média dos CCF é quase nulo. Assim, face ao reduzido número de observações e à instabilidade dos resultados por ano, poucas conclusões se podem retirar dos efeitos desta variável no CCF.

O «Indicador de Atividade» apresenta um padrão similar ao método de recolha variável. Neste sentido, o setor «Imobiliário» apresenta a maior média de CCF (de 44,1% considerando o CCF forçado ao intervalo unitário). Não obstante, o setor dos «Serviços», contrariamente ao método anterior, apresenta um CCF elevado relativamente à média da amostra (de 32,1%), com 39,8%. Por outro lado, o setor da «Manufatura», «Distribuição» e «Transportes» mantêm-se com as médias de CCF mais baixas (23,5%, 27,9% e 28,4% respetivamente). Tanto o setor «Financeiro», «Primário» e as «Entidades Públicas» têm poucas observações, pelo que pouco se poderá concluir quanto ao verdadeiro risco dos mesmos.



A «Colateralização» das linhas é também pouco relevante nesta amostra. As médias são também superiores nas linhas com colaterais no ano de 2012, onde atinge o máximo de 49,4% no CCF forçado ao intervalo unitário. Nas linhas sem colaterais, similarmente ao método de recolha anterior, a média foi mais alta no ano de 2011, com 34,7%. O mesmo sucede com o CCF forçado ao intervalo  $[-1,1]$ , sendo as médias sempre negativas com exceção dos contratos com colaterais cuja média no ano de 2012 é de 28,3% (claramente superior aos 6% do método anterior).

Sumariamente, neste método os resultados aproximam-se do método de recolha variável. Contudo, a correlação é mais elevada sobretudo no «Nível de Utilização» e no «Montante não utilizado», bem como na variável respeitante ao saldo de depósitos na IF. Ao nível das variáveis discretas, os resultados são similares, embora o número de observações limite mais severamente as conclusões. No que se refere aos métodos de tratamento de extremos, também neste método o CCF censurado ao intervalo unitário apresenta maior correlação com todas as variáveis, exceto o «Nível de utilização», o «Montante não utilizado» e o «Limite».

#### 4.3. RESULTADOS DOS MODELOS DE ESTIMAÇÃO

Os modelos desenvolvidos visam comparar os 2 métodos de recolha da amostra sobre diferentes perspectivas, isto é, considerando os aspetos do tratamento de extremos e as técnicas de estimação mais utilizadas. Seguidamente, detalham-se as opções tomadas quanto às técnicas de estimação utilizadas e os resultados de cada um destes métodos.

##### 4.3.1. MÉTODO DE ESTIMAÇÃO COM BASE EM MÉDIA DE CLASSES/GRUPOS

A média por grupos/classes é um dos métodos mais utilizados pelas IF, não só pela sua simplicidade de compreensão e implementação, mas sobretudo por ser robusto ao princípio de linearidade da relação entre o CCF e a variável independente. No desenvolvimento dos modelos optou-se por desenvolver várias alternativas de modelos que permitissem comparar diferentes opções tomadas pelas IF. Assim, considerou-se um modelo com 7 classes e outro com 5 classes por forma a analisar melhoria potencial ao introduzir mais grupos/classes. Por outro lado, analisou-se a qualidade de um modelo considerando apenas as variáveis relacionadas com a linha, isto é, o «limite», o «nível de utilização», o «montante utilizado» e o «montante não utilizado», dado que são as principais variáveis utilizadas pelas IF, uma vez que exigem menos desenvolvimentos tecnológicos na compilação de informação sobre os seus clientes. Por último, no sentido



de avaliar o impacto da utilização de CCF negativos, desenvolveu-se um modelo considerando o CCF forçado ao intervalo de  $[-1,1]$ . Previamente à seleção final das variáveis a integrar nos modelos, analisou-se a distribuição das médias de CCF por decis da variável que era selecionada automaticamente pela ferramenta *SAS Enterprise Miner*, para assim se compreender a relação entre o CCF e a variável. Seguidamente apresentam-se os principais resultados obtidos nos 2 métodos de recolha.

#### *Horizonte Temporal Variável*

Ao analisar a distribuição das variáveis por decis foi possível verificar que as principais variáveis selecionadas pela ferramenta *SAS Enterprise Miner* são consistentes com a análise univariada anterior. Assim, as variáveis «Nível de utilização» e «Montante não utilizado» apresentam a relação mais consistente com o CCF. No primeiro caso, as médias de CCF sobem à medida que o montante não utilizado aumenta, atingindo o máximo no último decil, com 42%, sendo que o 5.º decil (isto é, a mediana) está muito próxima da média desta amostra, com aproximadamente 25%. Por outro lado, o fato do valor da mediana do montante não utilizado ser de apenas 4200€, indica que variações nas estimativas de CCF nestes contratos terão um impacto reduzido no EAD, antecipando-se assim que as diferenças entre modelos serão pequenas. No segundo caso, a variável «nível de utilização», a média de CCF mais elevada verifica-se no 1.º decil, com 47% e onde o nível de utilização é de 58%. A relação é monotonicamente decrescente, isto é, o CCF decresce à medida que aumenta o nível de utilização. A variável «Rácio do saldo de depósitos sobre limite» apresenta também uma relação linear crescente, onde o CCF médio aumenta à medida que este rácio aumenta. Assim, quanto maior o saldo de depósitos face ao tamanho da linha, maior o CCF médio. A variável «Variação do limite face a 12 meses antes» tem uma relação menos linear. No 1.º decil, onde a variação é de 60%, ou seja, onde o limite foi reduzido até 60% comparativamente a 12 meses antes, o CCF médio é de 24%, reduzindo até ao 5.º decil, onde o corte foi de apenas 2% e a média de CCF é de 9%. No entanto, a partir do 6.º decil, onde não se registaram cortes de linha, ou onde estas inclusivamente foram aumentadas, o CCF médio era de 38%. Este comportamento sugere que nos contratos de «risco» onde as empresas sofrem cortes de linhas elevados, estas reagem utilizando o máximo disponível. Contrariamente, as empresas que não sofrem controlo tão apertado por parte da IF utilizam o remanescente das linhas antes de incumprirem, representando por isso o maior risco para as IF. Em

último, as variáveis «limite» e «Rácio do limite sobre ativo total», embora tenham sido selecionadas para os modelos não apresentam uma relação linear com o CCF.

No Anexo XIV encontram-se os resultados das opções de modelos referidas anteriormente. Em todos os modelos, é nas classes com maior CCF que se verifica o maior risco de subestimação do EAD (isto é, onde o desvio relativo do EAD estimado face ao EAD observado é mais baixo e em muitos casos negativo), sendo exceção o modelo de 5 classes, em que a média estimada mais elevada de 63% apresenta um desvio relativo positivo de aproximadamente 6%<sup>36</sup>. Por outro lado, o modelo baseado nas características da linha apresentou os maiores desvios relativos no EAD por classe. Contrariamente, o modelo com CCF forçado ao intervalo  $[-1,1]$  apresentou os piores resultados, subestimando o EAD em 2 classes. Tal como evidenciado em análises anteriores, as classes com as melhores pontuações, os valores de montantes não utilizados mais elevados e a maior proporção de depósitos sobre o limite da linha apresentaram as estimativas de CCF mais altas. No sentido inverso, a reduzida proporção dos depósitos face ao limite da linha, o controlo das linhas (i.e. a redução do limite), os baixos montantes não utilizados, o nível de utilização elevado e as piores pontuações apresentaram as estimativas de CCF mais baixas. Por outro lado, ao analisar os desvios relativos do EAD por mês ao longo do período de desenvolvimento verifica-se que, em média, todos os modelos tendem a compensar situações de subestimação, dado que em nenhum período se verificou uma média negativa deste indicador. Ainda assim, o modelo baseado no CCF forçado ao intervalo  $[-1,1]$  apresentou as médias mais baixas e o modelo baseado nas características da linha as mais altas. No que se refere à *performance* dos modelos, podemos verificar na Tabela 3, que embora o coeficiente de determinação varie dependendo do modelo, sendo mais elevado no CCF com *floor* de -100% (com 0.1755), o mesmo não sucede com o EAD estimado, onde o coeficiente é bastante semelhante entre os modelos. O desvio relativo médio no EAD é também mais baixo no CCF com *floor* de -100%, seguido pelo modelo com 5 classes. Contrariamente, o modelo baseado nas características da linha apresenta os maiores desvios relativos, aproximadamente 10%. Adicionalmente, conclui-se que a adição de mais classes não produz resultados significativos e as diferenças marginais que geram não justificam o aumento do nível de complexidade.

---

<sup>36</sup> Consultar Tabela 3 do Relatório.

Sumário da Performance dos Modelos				
Árvore	$R^2$ (CCF)	$R^2$ (EAD)	Erro quadrático médio (CCF)	Desvios Relativos - EAD
Árvores 7 Classes	0,1114	0,9049	0,147	8,5%
Árvores 5 Classes	0,0805	0,9040	0,153	8,4%
Árvores Características da Linha	0,0762	0,9053	0,153	9,6%
Árvores com <i>Floor</i> de -100%	0,1755	0,9052	0,477	6,0%

Tabela 3 – Sumário da *Performance* dos modelos de Árvores de Regressão – Amostra VTH restringida ao nível de utilização inferior a 99%.

### *Horizonte Temporal Fixo*

No método de horizonte fixo, as variáveis selecionadas são também consistentes com a análise univariada, embora com algumas diferenças face ao método anterior. Através da análise da distribuição das variáveis por decis constatou-se que as variáveis «Nível de utilização» e «Rácio do saldo de depósitos sobre limite» apresentaram a relação mais consistente com o CCF. Contrariamente ao método anterior, a variável «Montante não utilizado» não foi selecionada. No «nível de utilização» a média de CCF mais elevada verifica-se também no 1.º decil, com 66,5% e onde o nível de utilização é de 50%, decrescendo monotonicamente até ao nível de 99%. Na variável «Rácio do saldo de depósitos sobre limite» a relação linear crescente é mais acentuada comparativamente ao método anterior, sendo que no 5.º decil, atinge um CCF médio de 40,9% num rácio de 5,1%, bem acima da média da amostra (32,1%). A variável «Variação do saldo de créditos face a 12 meses antes» apresenta também uma relação muito semelhante à variável «Variação do limite face a 12 meses antes», onde o 1.º decil apresenta uma média mais elevada (de 33%) diminuindo até 6.º decil em que o montante de crédito não sofreu cortes. A partir do 7.º decil a média de CCF cresce rapidamente atingindo o máximo de 55% no último decil. Este comportamento sugere que as empresas se comportam estrategicamente, ou seja, procuram aumentar as suas fontes de financiamento quando percecionam que a IF não lhes atribui risco elevado, desta forma podem utilizá-las para liquidar atempadamente os montantes utilizados das linhas, garantindo que estas se mantenham disponíveis sem a intervenção da IF. Este comportamento é tanto mais visível se considerarmos um maior afastamento ao momento do incumprimento. A variável «Variação do limite face a 12 meses antes» apresenta um comportamento similar ao

método anterior, embora mais severo, no sentido em que as empresas que não sofrem cortes de linha apresentam uma média no CCF significativamente superior aos que sofrem cortes entre os 80% a 96%, ou seja, de 43% comparando com 7,3%. Por último, as variáveis «limite» e «Fornecedores de curto prazo» apresentam relações pouco claras, onde a média de CCF varia de forma inconstante ao longo da distribuição.

No Anexo XV podemos verificar que, tal como no método anterior, em todos os modelos, as classes com maior CCF têm maior risco de subestimação do EAD. Similarmente ao método anterior, os modelos baseados nas características da linha e apenas com 5 classes foram os que menos subestimaram o EAD por classe. O modelo baseado nas características da linha apenas utilizou 2 variáveis, o «nível de utilização» e o «limite», no entanto, apresentou resultados muito similares ao modelo de 5 classes, que utilizou as variáveis «nível de utilização», «variação saldo de créditos face a 12 meses antes» e «variação saldo de créditos face a 6 meses antes», indicando que a utilização de apenas as variáveis relacionadas com a linha pode gerar bons resultados. Neste método, o modelo com 7 classes apresentou os piores resultados, com maior subestimação por classes, contrariamente ao modelo baseado no CCF forçado ao intervalo  $[-1,1]$  que subestimou apenas marginalmente numa classe. Na análise dos desvios relativos do EAD por trimestre ao longo do período de desenvolvimento verifica-se que, em média, todos os modelos tendem a compensar situações de subestimação, dado que nenhum modelo apresentou uma média negativa deste indicador. Ainda assim, o modelo baseado no CCF forçado ao intervalo  $[-1,1]$  apresentou as médias mais baixas e o modelo baseado nas características da linha as mais altas. Quanto à *performance* dos modelos apresentados na Tabela 4 verifica-se que o coeficiente de determinação mantém-se também mais elevado no CCF com *floor* de -100% (com 0.2455), sendo também superior ao método anterior em todos os modelos. Contudo, se analisarmos o mesmo coeficiente considerando o EAD estimado verifica-se que em todos, este é inferior ao método anterior, embora a diferença não seja significativa. O coeficiente é também semelhante em todos os modelos, com exceção do modelo das características da linha, com um coeficiente ligeiramente inferior aos restantes (0.887). O desvio relativo médio no EAD é igualmente inferior no CCF com *floor* de -100%, seguido pelo modelo com 7 classes. Por outro lado, em todos os modelos o desvio relativo é superior comparativamente ao método anterior. O modelo baseado nas características da linha apresenta também os maiores desvios relativos, de 10,7%. Note-se que, tal como no método anterior, a introdução de mais classes não melhora os

resultados do modelo, dado que o coeficiente de determinação no EAD é inferior ao modelo com 5 classes.

Sumário da Performance dos Modelos				
Árvore	$R^2$ (CCF)	$R^2$ (EAD)	Erro quadrático médio (CCF)	Desvios Relativos – EAD
Árvores 7 Classes	0,1772	0,895	0,167	8,7%
Árvores 5 Classes	0,1869	0,896	0,157	10,4%
Árvores Características da Linha	0,1136	0,887	0,171	10,7%
Árvores com Floor de -100%	0,2455	0,895	0,510	6,9%

Tabela 4 – Sumário da *Performance* dos modelos de Árvores de Regressão – Amostra FTH restringida ao nível de utilização inferior a 99%.

Em suma, ambos os métodos apresentam resultados semelhantes na seleção das variáveis, sendo a criação das classes/grupos consistente entre os métodos. No entanto, o método de horizonte fixo apresenta um coeficiente de determinação mais baixo no EAD e os desvios relativos são superiores, embora este último facto possa representar menor risco de subestimação aquando da aplicação do modelo em momentos diferentes do tempo. Por outro lado, em ambos não se verifica uma melhoria significativa nos modelos pela utilização de outras variáveis alternativas às tipicamente utilizadas pelas IF, ou seja, as características da linha. A utilização de um CCF com valores negativos apresenta os menores desvios em ambos os métodos, embora sejam sempre positivos ao longo do período amostral. Ainda assim, podemos considerar mais conservadora a utilização de um CCF forçado ao intervalo unitário, uma vez que este permite uma reserva maior de capital ao atribuir um CCF estimado positivo para os contratos de menor risco que desta forma permitiam compensar a tendência dos modelos em subestimar o EAD nas classes com maiores CCF.

#### 4.3.2. MÉTODO DE ESTIMAÇÃO COM REGRESSÃO LINEAR MÚLTIPLA

No desenvolvimento dos modelos de regressão consideraram-se as variáveis descritas na análise univariada para cada tipologia de tratamento de extremos (i.e. o CCF forçado a  $[0,1]$  e a  $[-1,1]$ ) e em cada horizonte temporal de recolha da amostra. Foram desenvolvidos vários modelos alternativos, considerando, por um lado os efeitos da introdução de uma variável *dummy* para a melhor Pontuação («Pontuação 1») e, por outro,

os efeitos de redução da amostra aos contratos com nível de utilização inferior a 85%. Seguidamente apresentam-se os principais resultados dos modelos por horizonte temporal.

#### *Horizonte Temporal Variável*

Na Tabela 5 apresentam-se os resultados dos modelos de regressão da amostra baseada nos contratos com nível de utilização inferior a 99%. No que respeita ao CCF forçado ao intervalo unitário verifica-se que a variável *dummy* «Pontuação 1» permite melhorar o coeficiente da correlação do modelo face ao modelo sem esta variável (0,106 vs. 0,059). No entanto, ao nível do EAD estimado esta melhoria é quase nula (0,906 vs. 0,904) e os desvios relativos no EAD mantêm-se idênticos, próximos dos 10%. Nos modelos de CCF forçado ao intervalo  $[-1,1]$  verifica-se que a inclusão da variável «Pontuação 1» também melhora o modelo, mas não de forma tão significativa, sendo o coeficiente de correlação de 0.117 comparativamente a 0.091 sem esta variável. Em geral, verifica-se que o coeficiente de correlação no EAD mantém-se idêntico em todos, sendo aproximadamente de 0.90. Ainda assim, quanto aos desvios relativos no EAD denota-se que existe uma diferença relevante na forçagem do CCF, sendo que os desvios são mais elevados no CCF forçado ao intervalo unitário (aproximadamente 10% comparando com cerca de 6% no CCF forçado a  $[-1,1]$ ). Em suma, estas alternativas têm um impacto nulo nas estimativas de EAD devido ao elevado nível de utilização das linhas. Neste sentido, procurou-se desenvolver modelos que reduzissem esse efeito, com o intuito de verificar a existência de potenciais melhorias nos modelos.

	CCF forçado ao intervalo unitário		CCF forçado a $[-1,1]$	
Variáveis selecionadas	Modelo Com Pontuação	Modelo Sem Pontuação	Modelo Com Pontuação	Modelo Sem Pontuação
Constante	0,490105	0,556535	0,590610	0,619676
Pontuação 1 (variável <i>Dummy</i> )	0,241623		0,337054	
Saldo de depósitos em t	0,000001		0,0000004	0,000001
Nível de utilização	- 0,341145	- 0,379547	- 0,972310	- 0,956419
Rácio Depósitos/Créditos	- 0,028464	- 0,035497		0,000168
Variação no crédito face a 6 meses	0,000105	0,000100	0,000174	
Média empresa (variável <i>Dummy</i> )		0,047350		0,030603
Variação das dívidas totais face a 12 meses antes			0,032966	
Montante não utilizado				0,000001

Sumário da Performance dos Modelos				
Graus de Liberdade	5823	5823	6217	6217
RMSE	0,381	0,391	0,715	0,725
$R^2$ (CCF)	0,106	0,059	0,117	0,091
$R^2$ (EAD)	0,906	0,904	0,904	0,905
Desvios relativos no EAD	0,096	0,096	0,063	0,065

Tabela 5 – Coeficientes dos Modelos de regressão e Sumário da *Performance* dos Modelos – Amostra VTH restringida ao nível de utilização inferior a 99%.

Na Tabela 6 apresentam-se os resultados de modelos de uma amostra com nível de utilização inferior a 85%. Da sua análise destaca-se que embora o coeficiente de correlação no CCF seja inferior em todos os modelos (entre os 0.05 e os 0.01), o mesmo não sucede com o coeficiente de correlação do EAD, que melhora na amostra com níveis de utilização até 85%, aproximando-se dos 0.94. No que toca aos desvios relativos no EAD, verifica-se que a maior diferença encontra-se na inclusão da variável «Pontuação 1», cujos desvios estão próximos dos 10% (comparando com 8% nos modelos sem esta variável).

No que se refere à inclusão de variáveis nos modelos gerados pelas duas amostras, verifica-se que o «Nível de utilização» mantém-se relevante em todos, seguido pela variável «Variação no crédito face a 6 meses». Contrariamente, a variável «Rácio Depósitos/Créditos» parece ser mais relevante apenas nos modelos da amostra com nível de utilização inferior a 99% e a variável «Variação das dívidas totais face a 12 meses antes» nos modelos com CCF forçado ao intervalo  $[-1,1]$ .

	CCF forçado ao intervalo unitário		CCF forçado a $[-1,1]$	
Variáveis selecionadas	Modelo Com Pontuação	Modelo Sem Pontuação	Modelo Com Pontuação	Modelo Sem Pontuação
Constante	0,882327	0,707721	0,869998	0,635749
Pontuação 1 (variável <i>Dummy</i> )	0,210647		0,290577	
Saldo de depósitos em t				
Nível de utilização	- 0,743967	- 0,540788	-1,322559	- 1,054268
Rácio Depósitos/Créditos				
Variação no crédito face a 6 meses	0,000104		0,000183	0,000178
Variação das dívidas totais face a 12 meses antes			0,091304	0,095131
Saldo de depósitos dos sócios	0,000001	0,000002		
Montante não utilizado		0,000002		0,000002

RLE		7,57984E-08		
<b>Sumário da Performance dos Modelos</b>				
Graus de Liberdade	4095	4094	4345	4345
RMSE	0,371	0,377	0,743	0,749
$R^2$ (CCF)	0,052	0,022	0,030	0,014
$R^2$ (EAD)	0,939	0,939	0,938	0,939
Desvios relativos no EAD	0,108	0,088	0,092	0,077

Tabela 6 – Coeficientes dos Modelos de regressão e Sumário da *Performance* dos Modelos – Amostra VTH restringida ao nível de utilização inferior a 85%.

Da análise dos desvios relativos mensais<sup>37</sup>, em ambas as amostras (com nível de utilização inferior a 99% ou a 85%) conclui-se que as maiores diferenças entre os modelos encontram-se na forçagem do CCF, sendo os desvios relativos superiores quando consideramos o CCF forçado a  $[0,1]$ . Todavia, a diferença entre a consideração entre um *floor* de 0% ou de um *floor* negativo (neste caso, -100%), fica muito atenuada quando na amostra apenas se consideram as observações de contratos com nível de utilização inferior a 85%.

#### *Horizonte Temporal Fixo*

Tal como no método anterior, foram realizados vários modelos de regressão considerando uma amostra restringida ao nível de utilização de 99% e outra ao nível 85%. Na tabela 7 apresentam-se os resultados da primeira amostra, sobre os quais se destaca o efeito positivo da introdução da variável *dummy* «Pontuação 1», tendo melhorado os coeficientes de correlação do CCF forçado ao intervalo unitário e do CCF forçado ao intervalo  $[-1,1]$  (com um coeficiente de correlação de 0.163 e 0.201 respetivamente).

Variáveis selecionadas	CCF forçado ao intervalo unitário		CCF forçado a $[-1,1]$	
	Modelo Com Pontuação	Modelo Sem Pontuação	Modelo Com Pontuação	Modelo Sem Pontuação
Constante	0,492869	0,541520	0,524748	0,583536982
Pontuação 1 (variável <i>Dummy</i> )	0,227119		0,369835	
Nível de Utilização	- 0,538257	-0,588239	- 1,410673	-1,483797
Variação das dívidas totais face a 6 meses antes	0,200038	0,210203	0,349694	0,365350593

<sup>37</sup>Consultar em Anexos, o Anexo XVI e XVII.



Variação no montante utilizado face a 12 meses antes	0,018701	0,019628	0,041643	0,042923722
Limite		0,000000		4,53E-07
<b>Sumário da Performance dos Modelos</b>				
Graus de Liberdade	605	605	605	605
RMSE	0,401	0,410	0,766	0,777
$R^2$ (CCF)	0,163	0,126	0,201	0,179
$R^2$ (EAD)	0,896	0,896	0,885	0,885
Desvios relativos no EAD	0,035	0,042	0,001	0,010

Tabela 7 – Coeficientes dos Modelos de regressão e Sumário da *Performance* dos Modelos – Amostra FTH restringida ao nível de utilização inferior a 99%.

O mesmo efeito manteve-se na amostra com nível de utilização inferior a 85%, de acordo com a Tabela 8. Embora os modelos melhorem com a introdução desta variável, o seu efeito na variável final, o EAD, é nulo em todos os modelos gerados nas duas amostras. No que respeita às diferenças entre as amostras podemos concluir que, tal como no método de horizonte temporal variável, o coeficiente de correlação do EAD é superior na amostra com nível de utilização inferior a 85% (e.g. CCF forçado a  $[0,1]$ <sup>38</sup>: 0.896 vs. 0.936 e CCF forçado a  $[-1,1]$ : 0.885 vs. 0.937). Ainda assim, contrariamente ao método de recolha variável, os desvios relativos no EAD diferenciam-se significativamente entre as amostras com níveis de utilização distintos, sendo superiores na amostra com nível de utilização inferior a 85%, na qual se aproximam dos 8% no CCF forçado ao intervalo unitário, e dos 7% no CCF forçado  $[-1,1]$ . Na amostra de nível de utilização inferior a 99%, os desvios relativos são bastante mais baixos, sobretudo no CCF forçado a  $[-1,1]$ , estando próximos de 1%. Neste método, considerando a amostra restringida a 99%, os desvios relativos no CCF forçado a  $[0,1]$  são aproximadamente de 4%, significativamente mais baixos que no método anterior (com cerca de 10%).

Ao nível das variáveis incluídas nos modelos, contrariamente ao método anterior, a variável «Nível de Utilização» apenas é selecionada nos modelos da amostra com nível de utilização inferior a 99%. Por outro lado, a variável «Variação no montante utilizado face a 12 meses antes» é selecionada em todos os modelos das duas amostras, o que denota a relevância do comportamento de utilização da linha quando existe apenas uma

<sup>38</sup> Valores considerando os modelos com a inclusão da variável *dummy* «Pontuação 1».

observação por cada empresa. Neste método, a variação no montante das dívidas totais revelou-se importante em todos os modelos, embora com diferenças no momento de variação considerado. No que respeita à variação no crédito, que se revelou importante no método de horizonte variável, neste método apenas foi selecionada a variável «variação no crédito face a 12 meses antes» na amostra com nível de utilização inferior a 85%.

	CCF forçado ao intervalo unitário		CCF forçado a [-1,1]	
Variáveis selecionadas	Modelo Com Pontuação	Modelo Sem Pontuação	Modelo Com Pontuação	Modelo Sem Pontuação
Constante	- 0,077335	- 0,074997	- 0,981790	1,22900178
Pontuação 1 (variável <i>Dummy</i> )	0,189331		0,337263	
Nível de Utilização				-2,28759369
Variação das dívidas totais face a 6 meses antes	0,144946			0,267753541
Variação no montante utilizado face a 12 meses antes	0,030351	0,033712	0,054773	0,063102602
Variação no crédito face a 12 meses	0,086773	0,091456	0,186445	0,19243445
Variação das dívidas totais face a 12 meses antes		0,142792	0,260784963	
Montante não utilizado		0,000004		
<b>Sumário da Performance dos Modelos</b>				
Graus de Liberdade	403	403	403	403
RMSE	0,379	0,383	0,784	0,791
$R^2$ (CCF)	0,108	0,088	0,093	0,077
$R^2$ (EAD)	0,936	0,937	0,937	0,937
Desvios relativos no EAD	0,077	0,079	0,068	0,068

Tabela 8 – Coeficientes dos Modelos de regressão e Sumário da *Performance* dos Modelos – Amostra FTH restringida ao nível de utilização inferior a 85%.

Ao analisar os desvios relativos trimestrais<sup>39</sup>, em ambas as amostras (com nível de utilização inferior a 99% ou a 85%) conclui-se que existe maior instabilidade nos desvios comparativamente ao método de horizonte temporal variável, no entanto, mantém-se o mesmo padrão de comportamento ao nível das diferenças verificadas entre os modelos, sendo a forçagem do CCF a [0,1] a opção metodológica que tem maior impacto nas diferenças entre o EAD estimado. Por outro lado, neste método destaca-se que os

<sup>39</sup>Consultar em Anexos, o Anexo XVIII e XIX.

resultados são mais estáveis, embora com ligeiras diferenças no filtro aplicado ao nível de utilização das observações.

Em suma, ambos os métodos apresentam resultados similares ao nível do coeficiente de correlação do EAD, o que se deve ao elevado nível de utilização das linhas. Por outro lado, ao analisar as distribuições dos resíduos<sup>40</sup> de cada modelo verifica-se que em todos o pressuposto de normalidade não é verificado. Do mesmo modo, as distribuições dos resíduos *versus* valores estimados do CCF evidenciam um padrão não aleatório, resultado, em grande parte, da forçagem dos CCF observados. Ainda assim, de acordo com o teorema de Srivastava (1971), o elevado número de observações em todas as amostras (considerando que a amostra mais pequena contém 400 observações) permite concluir sobre a robustez estatística dos testes realizados e a consequente validade dos resultados obtidos. Assim, mantém-se válida a opção pelas IF de utilização de modelos de regressão múltipla para estimação do EAD/CCF, embora seja relevante para este portefólio a utilização do maior número de observações possível, que permita deste modo ultrapassar as limitações no pressuposto de normalidade dos resíduos, o qual se deve à forçagem dos CCF observados (note-se que neste portefólio a forçagem dos CCF representa cerca de 57% da amostra no método de horizonte variável).

## 5. VALIDAÇÃO DOS MODELOS

Na validação dos modelos foi utilizada uma amostra com informação entre janeiro e dezembro de 2014 que continha 12 observações por empresa. Nesta amostra, ao contrário da utilizada para o desenvolvimento dos modelos, não foram retirados registos com nível de utilização superior a 99% (ou a 85%), sendo que quando essa situação se verificava o CCF aplicado foi de 0%. No que a esta opção diz respeito, realça-se ainda que a mesma vai de encontro ao pressuposto regulamentar referido em capítulos anteriores, visto que a opção por um CCF diferente de zero em linhas onde o montante utilizado é superior ao limite (nível de utilização superior a 100%) iria provocar um efeito contrário ao pretendido (ou seja, uma diminuição da dívida). No sentido de testar o impacto, neste portefólio, das alternativas de modelização do CCF nas estimativas de EAD, adicionalmente à análise de performance, comparam-se os resultados dos modelos considerando o risco de subestimação e o seu grau de conservadorismo. Para o primeiro caso, utilizou-se o rácio dos erros gerados pela sobrestimação (RMSE) e os erros gerados

---

<sup>40</sup> Consultar em Anexos, o Anexo XX, XXI, XXII e XXIII

pelos casos de subestimação. No segundo caso, compararam-se os desvios relativos gerados com os desvios relativos obtidos se considerássemos um  $CCF=0\%$ . Por outro lado, procurou-se também analisar os modelos numa perspetiva de complexidade *versus* resultados marginais obtidos, tendo-se para o efeito considerado como comparativo base os resultados obtidos pela média simples da amostra de desenvolvimento. Seguidamente, apresentam-se os resultados dos testes efetuados com a amostra de validação.

### 5.1. ANÁLISE DE *PERFORMANCE* DOS MODELOS

Da análise dos resultados dos modelos de médias por classes/grupos, apresentados nas tabelas 9 e 10, podemos concluir que, em termos de *performance*, em ambos os métodos de horizonte temporal o coeficiente de correlação do CCF é superior nas árvores de decisão com o CCF forçado a  $[-1,1]$ . Inversamente, o pior resultado é dado pelas árvores construídas apenas com a informação sobre a linha. No método de horizonte temporal fixo os coeficientes de correlação no CCF são ligeiramente superiores ao método de horizonte variável nos modelos com 5 e 7 classes, mas inferiores no modelo com CCF forçado a  $[-1,1]$ .

Modelos	CCF Estimado	Desvios relativos – EAD	RMSE	RMSE ( $\overline{EAD} > EAD$ )(1)	RMSE ( $\overline{EAD} < EAD$ )(2)	Rácio $\frac{(1)}{(2)}$	$R^2(CC F)$	$R^2(EAD)$
Árvore com 7 Classes	7,80%	4,3%	36 537,7	35 324,0	31 997,5	1,1040	0,06570	0,94773
Árvore com 5 Classes	7,92%	4,4%	36 247,5	35 117,2	32 147,9	1,0924	0,06229	0,94855
Árvore com Características da Linha	7,53%	4,7%	36 869,3	35 871,4	31 653,7	1,1332	0,05247	0,94719
Árvore com <i>Floor</i> -100%	-3,70%	3,6%	35 917,2	34 233,3	33 369,7	1,0259	0,14124	0,94879

Tabela 9 – Sumário da *Performance* dos Modelos do método de Horizonte Temporal Variável com base nas médias por classes/grupos– Resultados na Amostra de Validação.

Modelos	CCF Estimado	Desvios relativos – EAD	RMSE	RMSE ( $\overline{EAD} > EAD$ )(1)	RMSE ( $\overline{EAD} < EAD$ )(2)	Rácio $\frac{(1)}{(2)}$	$R^2(CC F)$	$R^2(EAD)$
---------	--------------	-------------------------	------	------------------------------------	------------------------------------	-------------------------	-------------	------------

Árvore com 7 Classes	6,46%	4,2%	36 035,4	35 012,1	31 074,0	1,1267	0,09711	0,94873
Árvore com 5 Classes	9,02%	5,1%	36 219,0	35 730,3	30 129,8	1,1859	0,06898	0,94885
Árvore com Características da Linha	9,44%	5,3%	37 034,0	36 590,0	30 776,6	1,1889	0,05247	0,94719
Árvore com <i>Floor</i> -100%	-5,85%	4,3%	35 812,9	34 711,6	31 169,7	1,1136	0,10036	0,94941

Tabela 10 – Sumário da *Performance* dos Modelos do método de Horizonte Temporal Fixo com base nas médias por classes/grupos – Resultados na Amostra de Validação.

No entanto, no que respeita ao EAD, o método de horizonte fixo foi marginalmente superior em todos os modelos (próximo dos 0.95), com exceção do modelo baseado nas características da linha.

Nos modelos de regressão múltipla, os resultados apresentados nas tabelas 11 e 12, revelam que no horizonte variável os modelos, em geral, apresentam maior coeficiente de correlação no CCF na amostra restringida ao nível 99%. Inversamente, no horizonte fixo os modelos apresentaram melhor *performance* na amostra restringida a 85%. Todavia, no que concerne ao EAD as diferenças são marginais, onde todos os modelos apresentam um coeficiente próximo dos 0.95. Em ambos, a inclusão da variável *dummy* «Pontuação 1» tem impacto reduzido nos resultados, inferior ao que se verificou na amostra de desenvolvimento. A utilização de uma forçagem a -100% parece gerar os melhores resultados ao nível da *performance* do CCF. Em suma, neste portefólio os modelos baseados em árvores de decisão apresentaram resultados bastante similares aos modelos de regressão múltipla, podendo assim ser mais vantajosos dada a sua simplicidade de implementação e compreensão.

Modelos	CCF Estimado	Desvios relativos – EAD	RMSE	RMSE ( $\overline{EAD} > EAD$ )(1)	RMSE ( $\overline{EAD} < EAD$ )(2)	Rácio $\frac{(1)}{(2)}$	$R^2(\text{CCF})$	$R^2(\text{EAD})$
Nível de Utilização < 99%								
C/ Pontuação e CCF em [0,1]	7,2%	4,6%	36 208,7	35 542,1	29 705,00	1,1965	0,0782	0,9488
S/ Pontuação	7,0%	4,4%	36 415,0	35 223,4	31 295,88	1,1255	0,0505	0,9481

e CCF em [0,1]								
C/ Pontuação e CCF em [-1,1]	-4,1%	3,6%	35 864,6	34 453,2	31 698,87	1,0869	0,0815	0,9488
S/ Pontuação e CCF em [-1,1]	-4,8%	3,6%	36 001,2	34 450,9	31 734,00	1,0856	0,1110	0,9485
Nível de Utilização<85%								
C/ Pontuação e CCF em [0,1]	5,6%	5,4%	37 769,6	37 564,7	28687,99	1,3094	0,0553	0,9450
S/ Pontuação e CCF em [0,1]	6,0%	6,0%	40 791,6	41 211,7	28047,50	1,4693	0,0372	0,9373
C/ Pontuação e CCF em [-1,1]	2,4%	4,1%	36 256,3	35 359,8	30326,63	1,1659	0,0383	0,9478
S/ Pontuação e CCF em [-1,1]	2,6%	4,7%	37 383,6	37 050,4	28849,13	1,2842	0,0334	0,9455

Tabela 11 – Sumário da *Performance* dos Modelos do método de Horizonte Temporal Variável com base na Regressão múltipla – Resultados na Amostra de Validação.

Modelos	CCF Estimado	Desvios relativos - EAD	RMSE	RMSE ( $\widehat{EAD} > EAD$ ) (1)	RMSE ( $\widehat{EAD} < EAD$ ) (2)	Rácio $\frac{(1)}{(2)}$	$R^2(CC F)$	$R^2(EAD)$
Nível de Utilização<99%								
C/ Pontuação e CCF em [0,1]	7,7%	4,9%	35 943,9	35 309,9	29 665,6	1,1903	0,04476	0,94951
S/ Pontuação e CCF em [0,1]	8,3%	5,1%	36 741,4	36 198,6	29 425,6	1,2302	0,03088	0,94772
C/ Pontuação e CCF em [-1,1]	-7,9%	3,9%	36 036,5	34 819,4	31 321,6	1,1117	0,06984	0,94838
S/ Pontuação e CCF em [-1,1]	-6,9%	4,1%	36 340,6	35 209,9	30 775,1	1,1441	0,05790	0,94777
Nível de Utilização<85%								

C/ Pontuação e CCF em [0,1]	2,3%	3,7%	36 252,9	34 456,9	33 575,5	1,0263	0,09534	0,94803
S/ Pontuação e CCF em [0,1]	4,0%	5,6%	38 488,4	38 624,6	28 150,7	1,3721	0,08147	0,94382
C/ Pontuação e CCF em [-1,1]	-4,3%	2,8%	37 878,9	34 351,6	39 697,2	0,8653	0,09552	0,94286
S/ Pontuação e CCF em [-1,1]	3,3%	5,1%	38 111,6	37 850,2	29 510,8	1,2826	0,15417	0,94289

Tabela 12 – Sumário da *Performance* dos Modelos do método de Horizonte Temporal Fixo com base na Regressão múltipla – Resultados na Amostra de Validação.

Apesar dos modelos de CCF divergirem ligeiramente entre métodos diferentes e, sobretudo, quanto à técnica de forçagem de CCF utilizada, ao nível das estimativas de EAD, os modelos mostram-se muito idênticos entre si. Nas árvores de decisão todos os modelos apresentam um coeficiente de correlação aproximadamente de 0.95. Assim, embora variações nestes modelos tenham algum impacto na performance do CCF, na variável final, o EAD, o seu efeito é quase nulo. No mesmo sentido, o coeficiente de correlação do EAD não varia significativamente nos modelos de regressão múltipla, apresentando-se também próximo dos 0.95. Neste portefólio, o elevado nível de utilização (i.e., um montante não utilizado muito baixo) leva a que modelos complexos não se justifiquem, uma vez que, uma média simples permite obter resultados semelhantes.

De facto, de acordo com a tabela 13 e 14, se considerássemos a média da amostra de desenvolvimento (25,7% no método de horizonte variável e 32,1% no método de horizonte fixo) iríamos obter um coeficiente de correlação de 0.946 na amostra de validação em ambos os métodos. Adicionalmente, neste portefólio uma percentagem significativa das empresas diminui a exposição, pelo que considerar um CCF nulo é ainda assim conservador.

	Horizonte Temporal Variável	Horizonte Temporal Fixo
<b>Média CCF</b>	25,7%	32,1%
<b><math>R^2</math> (EAD)</b>	0,8931	0,8660
<b>Desvios Relativos –EAD</b>	0,0661	0,0618

Tabela 13 – Sumário da *Performance* da média simples nos dois métodos de horizonte temporal na Amostra de Desenvolvimento restringida ao nível de utilização inferior a 99%.

<b>Modelos</b>	<b>CCF Estimado</b>	<b>Desvios relativos - EAD</b>	<b>RMSE</b>	<b>RMSE (<math>\widehat{EAD} &gt; EAD</math>) (1)</b>	<b>RMSE (<math>\widehat{EAD} &lt; EAD</math>) (2)</b>	<b>Rácio <math>\frac{(1)}{(2)}</math></b>	<b><math>R^2(EAD)</math></b>
Horizonte Temporal Variável	7,6%	4,0%	37 079,4	35 159,8	34 485,9	1,01954	0,94610
Horizonte Temporal Fixo	9,5%	4,4%	37 018,6	35 491,5	33 189,0	1,0694	0,94651

Tabela 14 – Sumário da *Performance* da média simples nos dois métodos de horizonte temporal.

## 5.2. NÍVEL DE CONSERVADORISMO

No contexto regulamentar, o risco de subestimação do EAD deverá ser minimizado, procurando-se selecionar os modelos onde se verifiquem menos erros de subestimação. Neste sentido, com base nas tabelas anteriores, verifica-se que no método das médias por classes/grupos (ou árvores de decisão), os modelos com base na informação da linha apresentam o rácio de erros mais elevado, bem como o maior desvio relativo no EAD, o que sucede quer no método de horizonte variável (4,7%) quer no horizonte fixo (5,3%), embora neste último seja superior. Todavia, destaca-se que se verificou um menor risco de subestimação no método FTH, dado que o rácio entre os erros de sobrestimação sobre os erros de subestimação era superior neste método em todos os modelos.

Nos modelos de regressão este rácio é também superior a 1, sendo mais elevado na amostra restringida ao nível de utilização de 85% do método de horizonte variável, onde atinge o máximo de 1,469 (considerando o modelo com o CCF forçado a  $[0,1]$  e sem a variável «Pontuação 1»). Em todos os modelos, este rácio é mais elevado no CCF forçado ao intervalo unitário, sendo também os desvios relativos no EAD superiores. Nas duas alternativas de estimação, este rácio é superior a 1, sendo que os rácios mais baixos encontram-se ao nível dos modelos com CCF forçado a  $[-1,1]$  e na amostra de horizonte temporal fixo com nível de utilização inferior a 85%. Este último poderá estar relacionado com a menor estabilidade das estimativas. Por outro lado, se considerarmos a média simples da amostra de desenvolvimento, este rácio é ainda assim superior a 1 nos dois métodos de horizonte temporal, embora o horizonte fixo seja superior ao horizonte variável (1,069 vs. 1,019). No que respeita aos desvios relativos no EAD, as médias simples têm desvios bastante próximos dos modelos de árvores de decisão e de regressão múltipla (aproximadamente 4% nos dois métodos de horizonte temporal).



Na análise dos desvios relativos mensais no EAD para cada modelo, apresentados nos Anexo XXIV e XXV, incluiu-se ainda o desvio relativo mensal de um CCF nulo (i.e., considerar que o EAD será exatamente igual ao montante atualmente utilizado). Ao analisar estes desvios, verifica-se que caso se optasse por um CCF de 0%, os desvios seriam ainda assim positivos, ou seja, o EAD estimado seria superior ao observado. Desta forma, nestes portefólios o controlo das linhas por parte das IF garante que os acréscimos na exposição sejam quase nulos. Como referido anteriormente, as árvores de decisão construídas com a informação da linha têm os maiores desvios relativos em ambos os métodos de horizonte temporal. Contrariamente, a opção de um *floor* -100% apresenta os desvios mais baixos, ainda assim superiores a um CCF nulo. Contudo, denota-se que nas árvores de decisão o método de horizonte fixo apresenta os desvios mais altos em todos os modelos, distanciando-se com maior destaque da opção de um CCF nulo. Por outro lado, verifica-se também que o aumento da complexidade das árvores não melhora o modelo, dado que os desvios se aproximam da opção de um *floor* de -100%. Nos modelos de regressão múltipla os dois métodos apresentam comportamentos similares, sendo que a distância dos modelos à opção de um CCF nulo no método de horizonte fixo é também maior, mas inferior à verificada nas árvores de decisão. Por outro lado, verifica-se que, em ambos os métodos, a média simples apresenta os mesmos desvios relativos que um modelo de regressão com CCF forçado ao intervalo unitário (que como referido anteriormente tende a ser mais conservador que um CCF com um *floor* de -100%). Assim, poder-se-á concluir que a utilização de modelos mais complexos não aporta vantagens significativas quer na precisão das estimativas de EAD, quer na garantia de uma maior reserva de capital.

Em suma, o método de horizonte fixo revela-se mais conservador em ambas as alternativas de estimação do CCF, embora um número reduzido de observações possa limitar a estabilidade das estimativas, sobretudo nos modelos de regressão. Facto que poderá aumentar o risco de subestimação no contexto de uma amostra diferente. Ainda assim, a utilização do método de horizonte variável nos modelos de regressão gera bons resultados se considerarmos uma amostra com um nível de utilização inferior a 85%. Este método de horizonte temporal permite obter mais informação sobre o comportamento das empresas (até 12 observações mensais), pelo que a redução da amostra contendo apenas aquelas que maior risco representam para as IF (empresas com baixos níveis de utilização, face ao limite de crédito estabelecido) permite não só melhorar os modelos, mas também

garantir maiores reservas de capital se optarmos por atribuir um CCF nulo nas restantes empresas não incluídas no modelo. Note-se, ainda, que se optarmos por este nível de utilização, a inclusão de CCF negativos tem um impacto significativo na redução do risco de subestimação do EAD. No contexto da inclusão de CCF negativos, tal como referido por Valvonis (2008), é possível limitar os seus efeitos se considerarmos o máximo entre 0% e a estimativa de CCF gerada pelos modelos que incluem CCF observados negativos, tal como evidenciado no presente trabalho. Contudo, conclui-se que um *floor* de 0% no CCF observado mantém-se mais conservador em todos os modelos que a opção sugerida por aquele autor.

## 6. CONCLUSÕES

O presente trabalho permitiu concluir genericamente sobre os aspetos a seguir destacados:

- No portefólio de PME a atuação das IF na redução dos limites de crédito aprovados ou no bloqueio do acesso aos montantes da linha de crédito ainda disponíveis, tem um impacto relevante nas estimativas de CCF e, consequentemente, nas estimativas de EAD; na verdade, nestas situações há tendência para se gerarem valores extremos de CCF negativos;
- Constatou-se que este portefólio apresenta níveis de utilização das linhas de crédito muito elevados, face aos limites de crédito aprovados, comparativamente aos níveis médios de utilização de empresas de média e grande dimensão; na verdade, nesta situação há tendência para se gerarem valores extremos de CCF, negativos ou positivos consoante o EAD seja inferior ou não ao montante utilizado da linha no momento de referência;
- Verificou-se que o desenvolvimento de modelos complexos é usualmente pouco útil neste portefólio, já que a opção por modelos mais simples (p.e. médias por grupos/classes) permitiram alcançar níveis de *performance* similares e com igual robustez;
- As distribuições dos CCF obtidas pelas duas metodologias em apreço, FTH e VTH, têm ambas tendência para maior concentração em valores negativos; assim, a utilização destas amostras em modelos de regressão múltipla obrigou à forçagem de dados, com vista a que as estimativas de CCF fossem positivas (i.e.

estabelecimento de *floors* para o CCF observado), o que originou a não normalidade dos resíduos.

Por outro lado, sobre o objetivo específico de se avaliar a utilização de diferentes metodologias de recolha da amostra para estimar o parâmetro EAD/CCF, com enfoque nos métodos FTH e VTH, concluiu-se o seguinte:

- A opção metodológica de recolha das amostras (i.e. entre FTH e VTH) teve um impacto reduzido nas estimativas de EAD, embora se constatem algumas diferenças descritas nos pontos seguintes;
- O método FTH revelou-se mais adequado nos modelos de EAD/CCF obtidos através de médias por classes/grupos, dado o menor risco de subestimação do EAD em todos os modelos (i.e. maior conservadorismo da estimativa de EAD);
- Por sua vez, tendo em conta as características da distribuição dos CCF observados, o método VTH revelou-se mais adequado, em termos conservadorismo de EAD, aos modelos de EAD/CCF obtidos a partir de modelos de regressão múltipla nas amostras com níveis de utilização mais baixos (p.e. inferior a 85%); apesar da não normalidade dos resíduos, a robustez estatística dos resultados ficou assegurada pelo elevado número de observações que esta metodologia proporciona (v. teorema de Strivastava);
- Com isto não se pretende afastar a utilização do método FTH na obtenção de modelos através de regressão múltipla, desde que o número de observações na amostra seja suficientemente amplo para garantir as condições de robustez da mesma, proporcionadas através da aplicação do teorema de Strivastava;
- No contexto das opções de tratamento dos valores extremos (*floors* e *caps*) verificou-se que em ambas as metodologias de recolha da amostra (FTH e VTH), a utilização de um *Floor* de 0% nos CCF observados utilizados no modelo mantém-se mais conservadora comparativamente à opção de um *Floor* de 0% nas estimativas de CCF resultantes do próprio modelo (i.e. nesta segunda opção, utilizando-se CCF negativos no modelo);
- A opção metodológica de recolha das amostras (FTH e VTH) não teve impacto ao nível das variáveis explicativas mais relevantes neste portefólio, mais especificamente o «Nível de utilização» da linha de crédito e a «Pontuação» ou *rating* da empresa, já que em ambas as metodologias aquelas variáveis foram as mais relevantes na capacidade explicativa do CCF; estes resultados foram também

consistentes com os de outros estudos realizados sobre os portfólios de empresas de média e grande dimensão;

- O histórico da relação da IF com a empresa é também relevante neste portfólio, uma vez que as variáveis essencialmente ligadas à relação da IF com a empresa (p.e. variações nos saldos de depósitos e créditos da empresa) foram também selecionadas seguidamente às duas variáveis mais importantes, referidas no ponto anterior;
- Ao contrário, verificou-se que as variáveis relativas à informação financeira disponibilizada pela empresa, o «limite», a «dimensão da empresa», e a «colateralização» são pouco significativas;

Neste portfólio, conclui-se que em ambas as metodologias o aumento da complexidade dos modelos, utilizando a regressão múltipla, não permite, por um lado, aumentar a *performance* dos modelos, e por outro, garantir maior grau de conservadorismo, pelo que poderá ser mais adequada a utilização de modelos mais simples, como as médias por classes ou grupos. Neste contexto, a metodologia FTH revela-se mais adequada, tendo em consideração que, para a mesma *performance*, é possível obter maior grau de conservadorismo nas estimativas de EAD.

Em suma, as conclusões do presente trabalho são consistentes com os principais eixos orientadores das entidades de supervisão nesta matéria, através das propostas de nova legislação em fase de consulta<sup>41</sup>, na seguinte medida: (a) utilidade de se garantir um *floor* nas estimativas de EAD, que tenha como mínimo o montante atualmente utilizado (i.e. um *floor* de 0% no CCF estimado), evitando-se estimativas de CCF negativas; (b) o interesse de se evitar o enviesamento da estimativa de CCF originado pela denominada «região de instabilidade», associada a linhas de crédito com níveis de utilização muito elevados na data de referência; assim, as IF poderão optar por filtrar as observações com níveis de utilizações das linhas de crédito próximos dos limites de crédito aprovados; (c) a clara preferência do regulador pelo método horizonte fixo (12.º mês anterior ao incumprimento), vai ao encontro dos resultados aqui apresentados, dado que se concluiu que este método proporciona um maior grau de conservadorismo, quando associado à modelização através de médias por grupos/classes.

---

<sup>41</sup> *Reducing variation in credit risk-weighted assets – constraints on the use of internal model approaches*, Consultative Document.

## 7. BIBLIOGRAFIA

Basel Committee on Banking Supervision (2005), *An Explanatory Note on the Basel II IRB Risk Weight Functions*, Working Paper.

Basel Committee on Banking Supervision (2016), *Reducing variation in credit risk-weighted assets – constraints on the use of internal model approaches*, Consultative Document.

Jimenez, Gabriel, Lopez, Jose A. e Jesus Saurina, *EAD Calibration for Corporate Credit Lines*, Banco de Espanha e Banco da Reserva Federal de São Francisco, Working Paper, 2008;

Moral, Gregorio, “*The Basel II Risk Parameters: Estimation, Validation and Stress Testing*”, Capítulo XI - *EAD Estimates for Facilities with Explicit Limits*, 2006;

Sufi, Amir, *Bank Lines of Credit in Corporate Finance: An Empirical Analysis*, University of Chicago, 2005;

Araten, Michel e Jacobs Jr, Michael, *Loan Equivalents for Revolving Credits and Advised Lines*, The RMA Journal, Abril 2001, pag. 34-39;

European Banking Authority, *Report on the Comparability of Supervisory Rules and Practices*, 17 dezembro 2013;

European Banking Authority, *Consultation Paper on Draft RTS on Assessment Methodology for IRB Approach*, 12 de novembro de 2014;

European Banking Authority, *Discussion Paper and Call for Evidence on SMEs and the SME Supporting Factor*, 31 de julho de 2015;

Regulamento Europeu N.º 575/2013 do Parlamento Europeu e do Conselho 26 de junho de 2013.

Berger, Allen N. e Udell, Gregory F., *Relationship Lending and Lines of Credit in Small Firm Finance*, New York University, Working Paper, 1994.

Valvionis, Vytautas, *Estimating EAD for retail exposures for Basel II purposes*, Journal of Credit Risk, Vol. 4, nº1, Spring 2008, pag. 79-109.

Strivastava, Muni e Sem, Ashish, *Regression Analysis: Theory, Methods, and Applications*, Springer, pag. 106-107.

Jacobs, Michael Jr, *An Empirical Study of Exposure at Default*, OCC Working Paper, Washington, DC: Office of the Controller of the Currency, 2009.

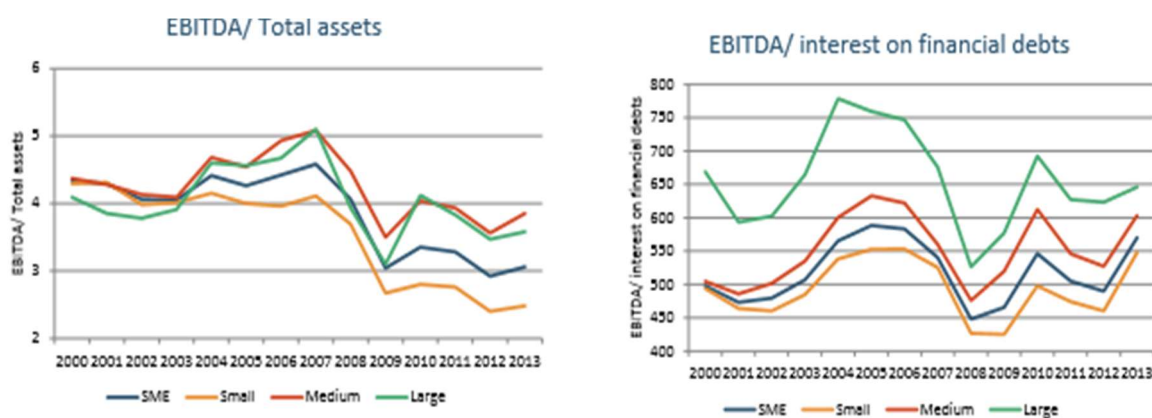
Brown, Iain L. J., PhD, *Developing Credit Risk Models Using SAS Enterprise Miner and SAS/STAT: Theory and Applications*, 2014

Barakova, Irina e Parthasarathy, Harini, *Moeling Corporate Exposure at Default*, Office of the Controller of the Currency, março 2013;

Afifi, A. A. e Clark, Virginia, *Computer- Aided Multivariate Analysis*, Lifetime Learning Publications Belmont, California, 1984

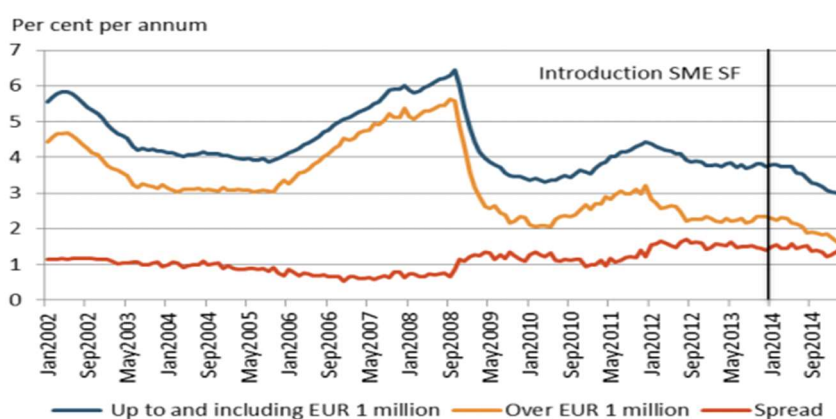
## 8. ANEXOS

### *Anexo I – Indicadores de empresas não financeiras na Zona Euro*



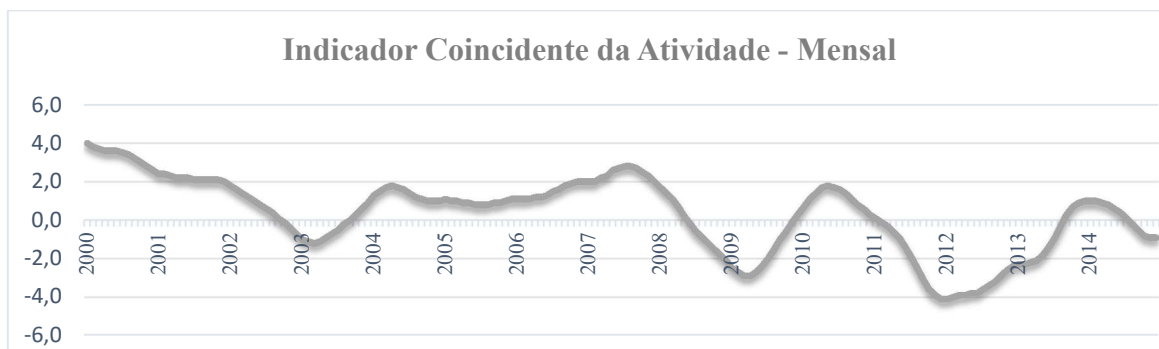
Fonte: *Discussion Paper and Call For Evidence on SMEs and SME Supporting Factor*, EBA; página 25.

### *Anexo II– Taxas de Juro médias de empresas não financeiras na Zona Euro*



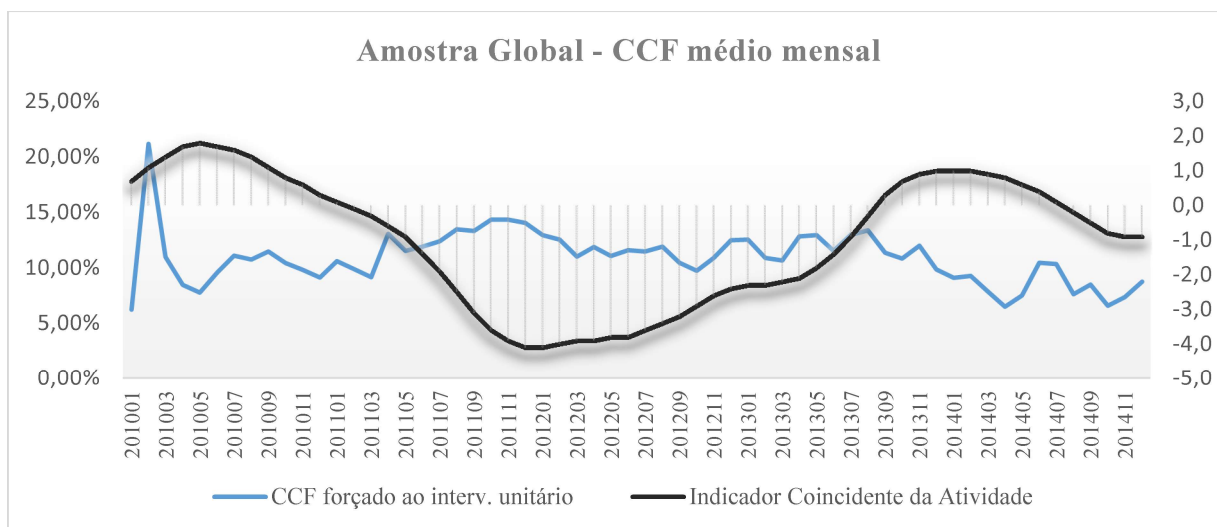
Fonte: *Discussion Paper and Call For Evidence on SMEs and SME Supporting Factor*, EBA; página 34.

*Anexo III – Evolução mensal do Indicador Coincidente da Atividade entre 2000 e 2014*



Fonte: Dados disponíveis no portal do Banco de Portugal, em <http://www.bportugal.pt/pt-pt/estudoseconomicos/publicacoes/indicadoresconjuntura/Paginas/IndicadoresdeConjuntura.aspx>

*Anexo IV – Evolução média mensal do CCF forçado ao intervalo unitário na Amostra global (período de 2010 a 2014)*



*Anexo V – Análise dos CCF médios por tipologia de forçamento do CCF, por Pontuação e Nível de utilização na Amostra VTH após filtragem ao nível de utilização inferior a 99%*

Pontuação	Nível de Utilização	CCF médio	CCF médio forçado a $[-1, 1]$	CCF médio forçado a $[0, 1]$
1	$0 \leq \text{Util} < 60\%$	63,9%	63,3%	67,4%
1	$0.6 \leq \text{Util} < 70\%$	21,9%	23,4%	45,2%
1	$0.7 \leq \text{Util} < 85\%$	34,5%	44,9%	59,0%
1	$0.85 \leq \text{Util} < 90\%$	-75,8%	-2,5%	38,6%



Avaliação do Impacto do Horizonte Temporal nas Estimativas de EAD de Linhas de Crédito com Limites  
Explícitos a PME

1	0.9=<Util<95%	-52,2%	8,3%	46,5%
	0.95=<Util<99%	-325,1%	-4,4%	36,4%
	<b>Subtotal</b>	<b>-102,5%</b>	<b>20,5%</b>	<b>48,4%</b>
2	0=<Util<60%	45,3%	44,6%	51,8%
2	0.6=<Util<70%	5,0%	9,1%	32,5%
2	0.7=<Util<85%	-19,3%	3,1%	37,5%
2	0.85=<Util<90%	-93,9%	-15,2%	32,8%
2	0.9=<Util<95%	-126,6%	-19,4%	28,9%
	0.95=<Util<99%	-326,9%	-24,7%	26,7%
	<b>Subtotal</b>	<b>-160,6%</b>	<b>-8,7%</b>	<b>32,6%</b>
3	0=<Util<60%	23,1%	21,5%	29,6%
3	0.6=<Util<70%	-15,3%	-12,8%	18,8%
3	0.7=<Util<85%	-34,7%	-16,8%	19,1%
3	0.85=<Util<90%	-62,9%	-27,6%	17,4%
3	0.9=<Util<95%	-107,7%	-24,7%	17,2%
	0.95=<Util<99%	-267,5%	-38,6%	12,4%
	<b>Subtotal</b>	<b>-154,3%</b>	<b>-26,2%</b>	<b>16,3%</b>
Sem Pontuação	0=<Util<60%	5,7%	7,1%	30,7%
Sem Pontuação	0.6=<Util<70%	-47,7%	-28,9%	13,3%
Sem Pontuação	0.7=<Util<85%	-130,3%	-54,0%	0,0%
Sem Pontuação	0.85=<Util<90%	-14,5%	-10,6%	15,3%
Sem Pontuação	0.9=<Util<95%	-161,2%	-26,1%	19,6%
Sem Pontuação	0.95=<Util<99%	-175,0%	-30,5%	6,6%
	<b>Subtotal</b>	<b>-131%</b>	<b>-26%</b>	<b>12%</b>

*Anexo VI – Análise dos CCF médios por tipologia de forçamento do CCF, por Pontuação e Nível de utilização na Amostra FTHapós filtragem ao nível de utilização inferior a 99%*

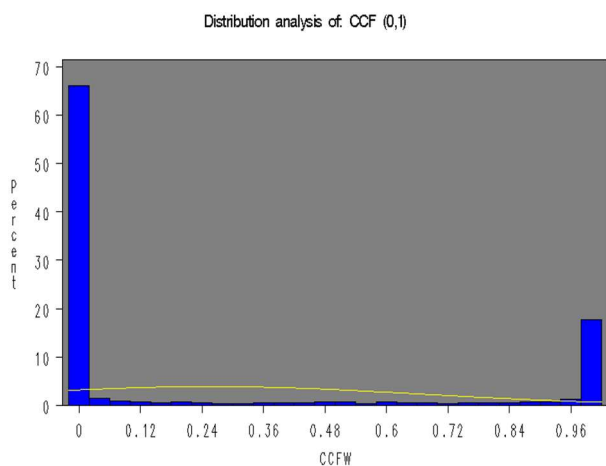
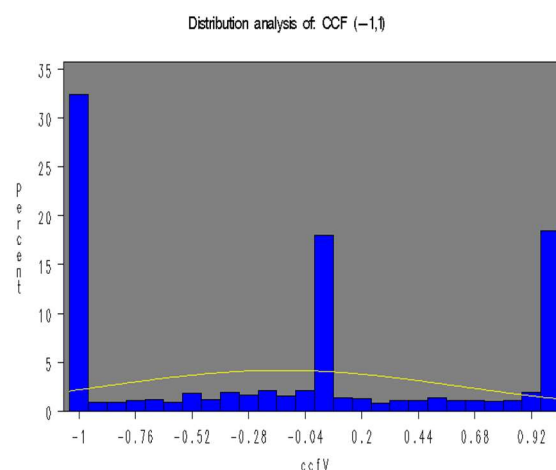
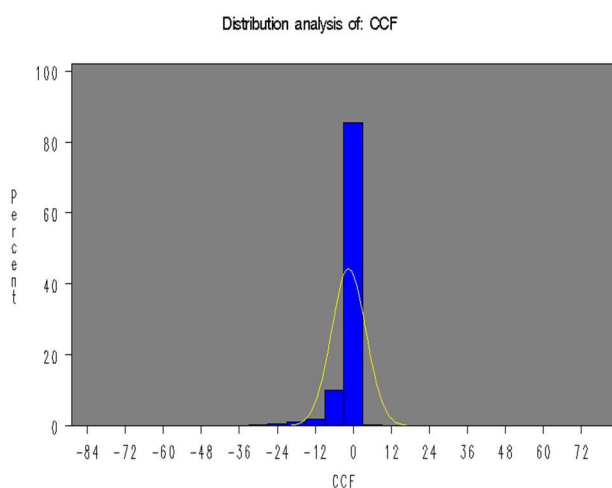
Pontuação	Nível de Utilização	CCF médio	CCF médio forçado a [-1, 1]	CCF médio forçado a [0, 1]
1	0=<Util<60%	62,8%	63,3%	68,0%
1	0.6=<Util<70%	49,7%	39,4%	66,2%
1	0.7=<Util<85%	49,4%	64,6%	76,3%
1	0.85=<Util<90%	-224,8%	-30,0%	30,0%
1	0.9=<Util<95%	-28,7%	-1,6%	42,8%
	0.95=<Util<99%	-542,3%	-5,8%	41,3%
	<b>Subtotal</b>	<b>-155,0%</b>	<b>23,3%</b>	<b>54,1%</b>
2	0=<Util<60%	56,0%	56,4%	67,5%
2	0.6=<Util<70%	-1,3%	-1,3%	27,6%
2	0.7=<Util<85%	-38,4%	-2,1%	38,4%
2	0.85=<Util<90%	-73,1%	-27,0%	25,3%
2	0.9=<Util<95%	-145,5%	-28,4%	28,7%
	0.95=<Util<99%	-443,1%	-34,1%	29,2%
	<b>Subtotal</b>	<b>-201,1%</b>	<b>-14,2%</b>	<b>34,7%</b>
3	0=<Util<60%	43,1%	43,1%	47,4%
3	0.6=<Util<70%	-45,0%	-33,1%	18,4%
3	0.7=<Util<85%	-21,2%	6,1%	35,7%
3	0.85=<Util<90%	-109,4%	-32,3%	19,2%
3	0.9=<Util<95%	-182,7%	-34,1%	21,7%
	0.95=<Util<99%	-388,9%	-70,1%	7,5%
	<b>Subtotal</b>	<b>-220,4%</b>	<b>-37,2%</b>	<b>19,4%</b>
Sem Pontuação	0=<Util<60%	36,5%	36,5%	36,6%
Sem Pontuação	0.6=<Util<70%	0,1%	0,1%	0,1%
Sem Pontuação	0.7=<Util<85%	-	-	-

Avaliação do Impacto do Horizonte Temporal nas Estimativas de EAD de Linhas de Crédito com Limites  
Explícitos a PME

Sem Pontuação	0.85=<Util<90%	100,3%	100,0%	100,0%
Sem Pontuação	0.9=<Util<95%	-165,0%	-7,7%	42,3%
Sem Pontuação	0.95=<Util<99%	-507,2%	-62,5%	0,0%
<b>Subtotal</b>		<b>-241,9%</b>	<b>-16,8%</b>	<b>23,2%</b>

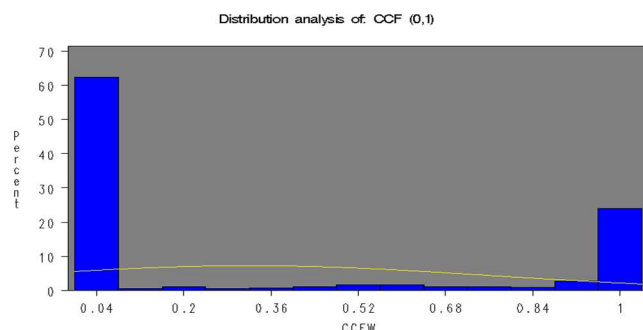
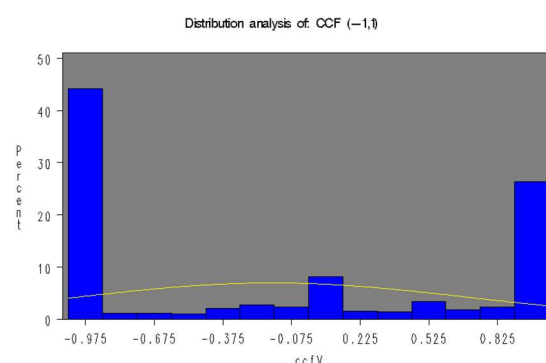
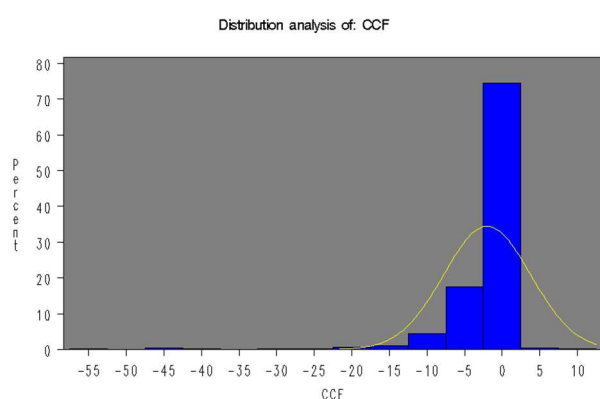
*Anexo VII – Sumário estatístico da Amostra VTH restringida ao nível de utilização inferior a 99%*

Variável	#	Média	Desvio-padrão	Mínimo	5.º Pctl	25.º Pctl	50.º Pctl	75.º Pctl	95.º Pctl	Máximo	Assimetria	Curtose
<b>CCF</b>	6223	-146,6%	6,8%	-8410,6%	792,4 %	145,0 %	0,0 %	53,7 %	100,0 %	7920,9 %	-5,783	69,505
<b>CCF forçado a [-1, 1]</b>	6223	-13,9%	1,0%	-100,0%	100,0 %	100,0 %	0,0 %	53,7 %	100,0 %	100,0%	0,253	-1,353
<b>CCF forçado a [0, 1]</b>	6223	25,7%	0,5%	0,0%	0,0%	0,0%	0,0 %	53,7 %	100,0 %	100,0%	1,105	-0,634



*Anexo VIII – Sumário estatístico da Amostra FTH restringida ao nível de utilização inferior a 99%*

Variável	#	Média	Desvio-padrão	Mínimo	5.º Pctl	25.º Pctl	50.º Pctl	75.º Pctl	95.º Pctl	Máximo	Assimetria	Curtose
CCF	682	200,4%	22,2%	5726,3%	958,6%	250,0%	23,5%	94,0%	100,0%	1233,5%	-5,314	37,93
CCF forçado a [-1, 1]	682	-15,8%	3,3%	100,0%	100,0%	100,0%	23,5%	94,0%	100,0%	100,0%	0,288	-1,65
CCF forçado a [0, 1]	682	32,1%	1,7%	0,0%	0,0%	0,0%	0,0%	94,0%	100,0%	100,0%	0,746	-1,335

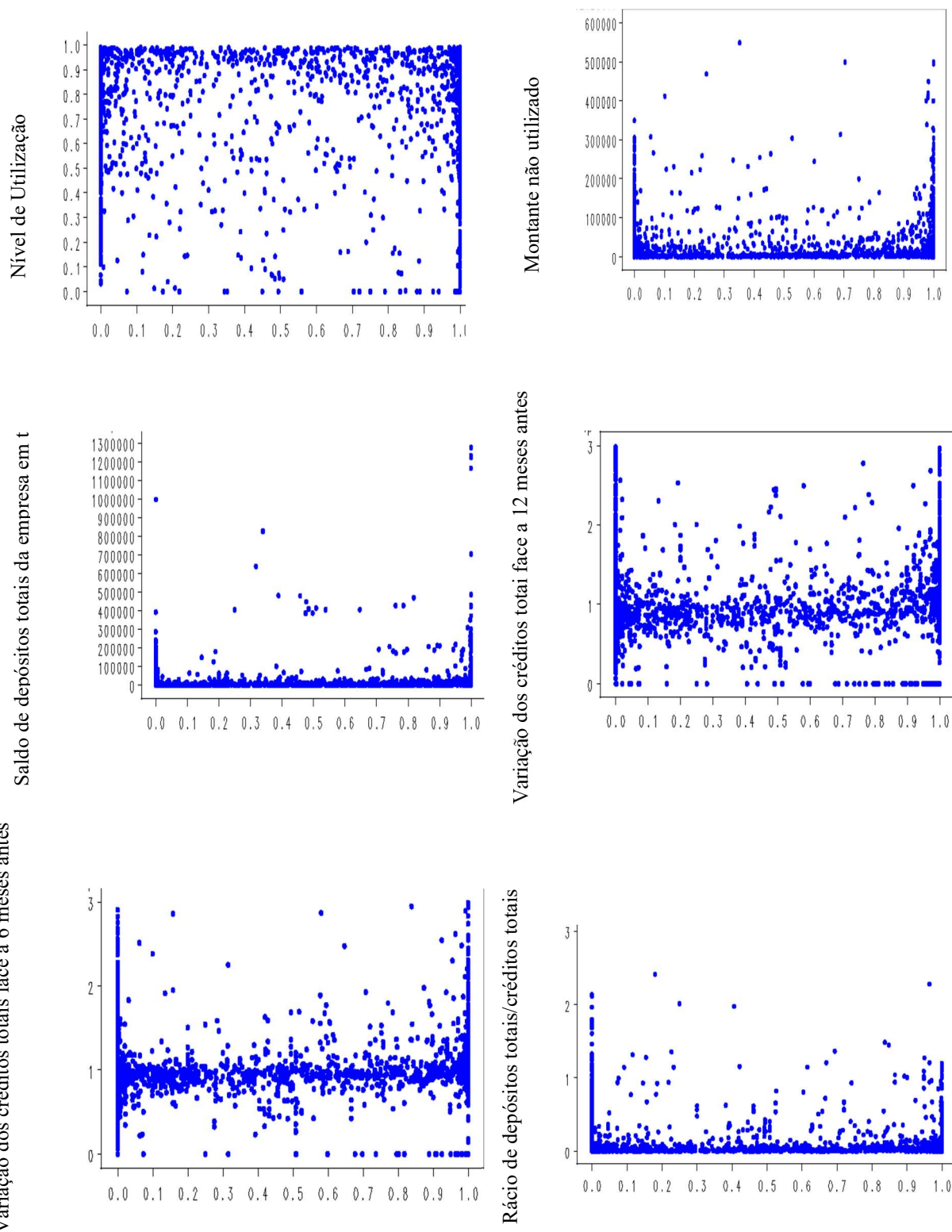


*Anexo IX – Descrição das variáveis utilizadas no desenvolvimento dos modelos em ambos os métodos de recolha*

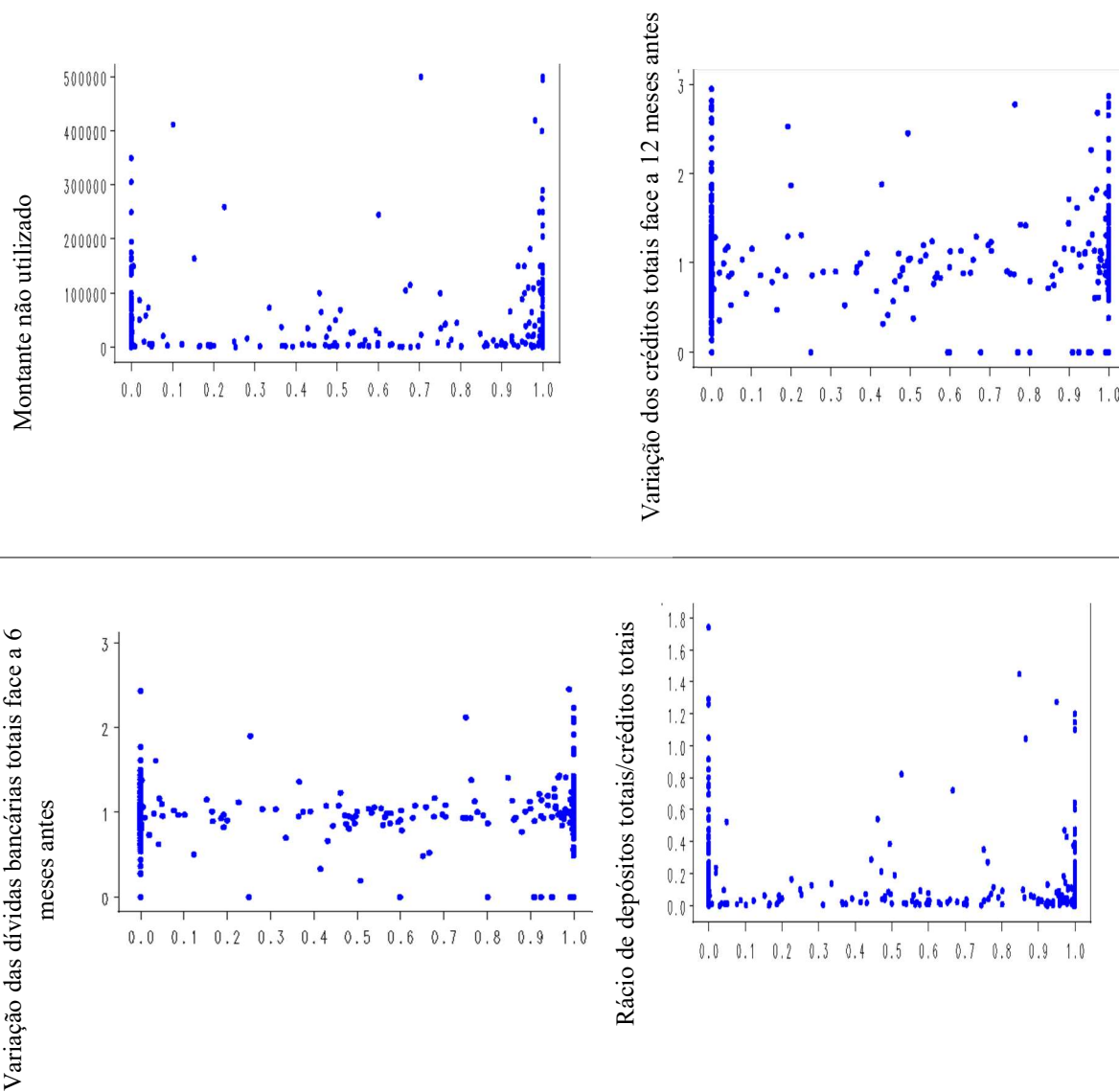
Categoria	Variáveis	Descrição
Características da Linha	Nível de Utilização	$\frac{\text{Montante Utilizado}_t}{\text{Limite}_t}$
	Montante não utilizado	$\text{Limite}_t - \text{Montante Utilizado}_t$
	Limite	Limite contratado (considerando alterações contratuais ao longo do período de observação de 12 meses)
	Variação no Montante Utilizado	$\frac{\text{Montante Utilizado}_t}{\text{Montante Utilizado}_{t-12}}$

	Variação no Limite	$\frac{Limite_t}{Limite_{t-12}}$
<i>Rating</i>	Pontuação	O <i>Rating</i> foi reclassificado em 3 pontuações, onde 1=A, B; 2=C,D e 3=E,F. Esta reclassificação considerou a maior correlação com o CCF forçado ao intervalo unitário nos 2 métodos de recolha.
Informação de outros contratos com a IF	Saldo de créditos totais da empresa em t	Saldo de todos os créditos que a empresa obteve na instituição na data de referência (t).
	Saldo de depósitos totais dos sócios em t	Saldo de todos os depósitos que os sócios da empresa detém na instituição na data de referência (t).
	Saldo de depósitos totais da empresa em t	Saldo de todos os depósitos que a empresa detém na instituição na data de referência (t).
	Variação do saldo de créditos totais face a 12 meses antes	$\frac{Saldo\ de\ Créditos\ Totais\ da\ empresa_t}{Saldo\ de\ Créditos\ Totais\ da\ empresa_{t-12}}$
	Variação do saldo de créditos totais face a 6 meses antes	$\frac{Saldo\ de\ Créditos\ Totais\ da\ empresa_t}{Saldo\ de\ Créditos\ Totais\ da\ empresa_{t-6}}$
	Rácio depósitos totais em t/créditos totais em t	$\frac{Saldo\ de\ Depósitos\ Totais\ da\ empresa_t}{Saldo\ de\ Créditos\ Totais\ da\ empresa_t}$
Informação da Central de Responsabilidade de Crédito (Banco de Portugal)	Variação da dívida bancária total face a 6 meses antes	$\frac{Saldo\ da\ dívida\ total\ em\ IF_t}{Saldo\ da\ dívida\ total\ em\ IF_{t-6}}$
	Variação da dívida bancária total face a 12 meses antes	$\frac{Saldo\ da\ dívida\ total\ em\ IF_t}{Saldo\ da\ dívida\ total\ em\ IF_{t-12}}$
	Crédito Vencido à 1 Mês	Crédito em atraso à 1 mês em alguma instituição de crédito em Portugal
Informação do Balanço referente a 2 anos antes da data de referência (t-24)	<i>Return on Equity</i> em t-24	$\frac{Resultado\ líquido\ do\ exercício\ (RLE)_{t-24}}{Capital\ Próprio_{t-24}}$
	Indicador de Atividade	A informação do CAE foi agrupada em 8 categorias: entidades públicas;serviços;atividades financeiras; indústria;transportes; distribuição; imobiliário; setor primário.
	Resultado Líquido do Exercício em t-24	Resultado apresentado no balanço mais recente (pelo menos 24 meses antes da data de referência)
	Caixa e depósitos bancários em t-24	Numerário e depósitos apresentados no balanço mais recente.
	Rácio dívidas bancárias em t-24/Passivo total em t-24	$\frac{Dívidas\ de\ médio\ e\ longo\ prazo\ em\ IF_{t-24}}{Passivo\ Total_{t-24}}$
	Resultados Financeiros em t-24	Saldo dos custos e receitas financeiras apresentados no balanço mais recente (pelo menos 24 meses antes da data de referência).

*Anexo X – Análise gráfica da correlação das 6 principais variáveis explicativas com o CCF forçado ao intervalo  $[0,1]$  – Amostra VTH restringida ao nível de utilização inferior a 99%*

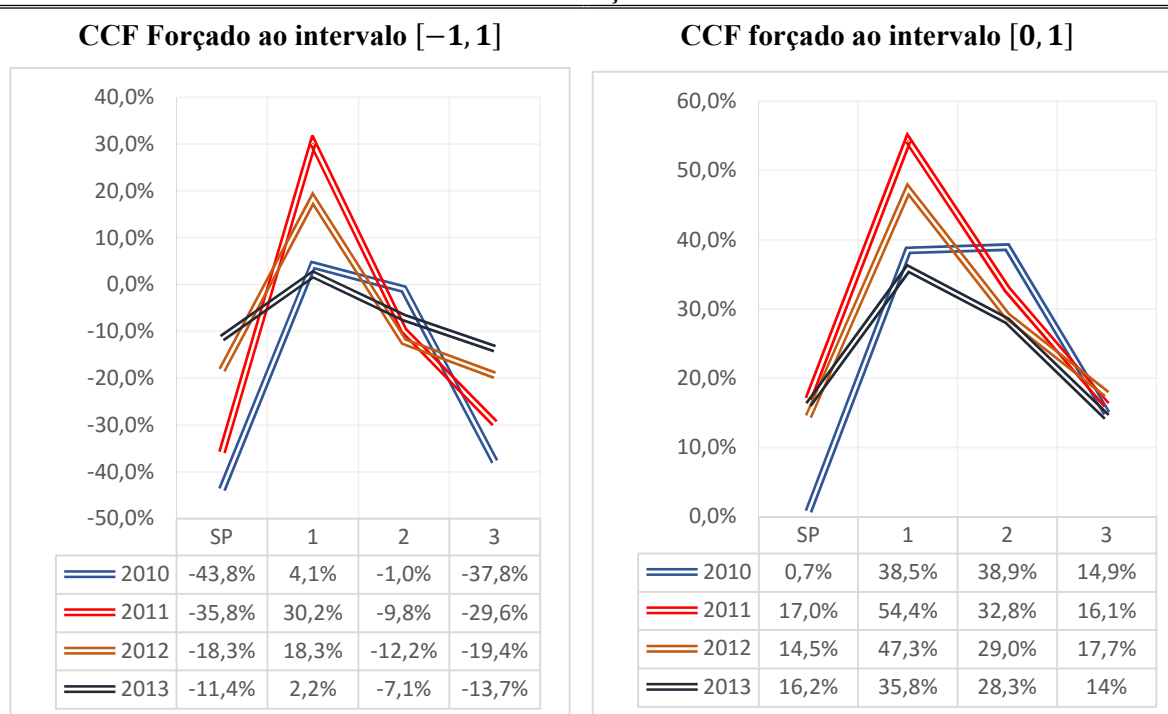


*Anexo XI – Análise gráfica da correlação das 6 principais variáveis explicativas com o CCF forçado ao intervalo  $[0,1]$  - Amostra FTH restringida ao nível de utilização inferior a 99%*

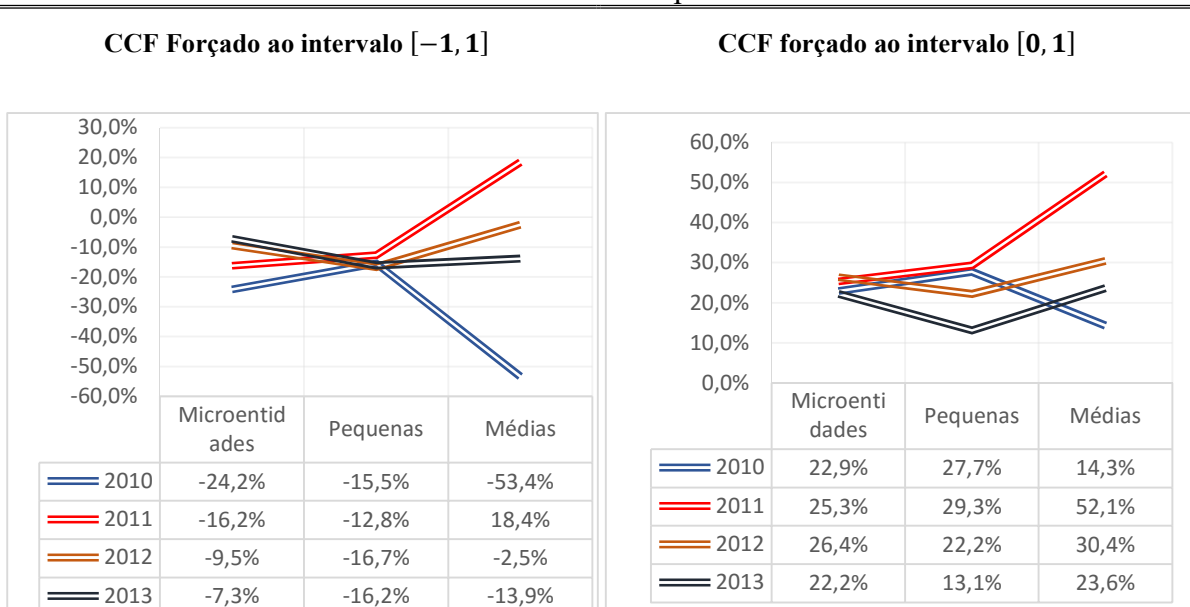


*Anexo XII – Análise gráfica das médias do CCF forçado ao intervalo  $[-1,1]$  e do CCF forçado ao intervalo  $[0,1]$  das Variáveis discretas Pontuação, Dimensão da Empresa, Colaterais e Indicador de Atividade - Amostra de VTH restringida ao nível de utilização inferior a 99%.*

### Pontuação



### Dimensão da Empresa



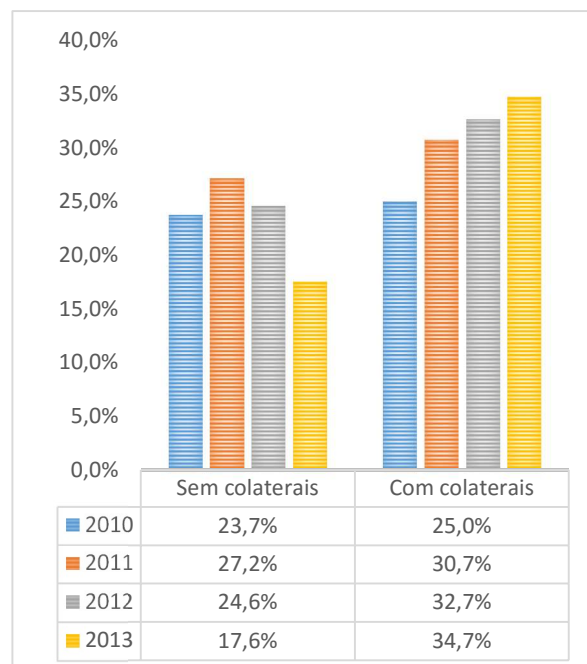
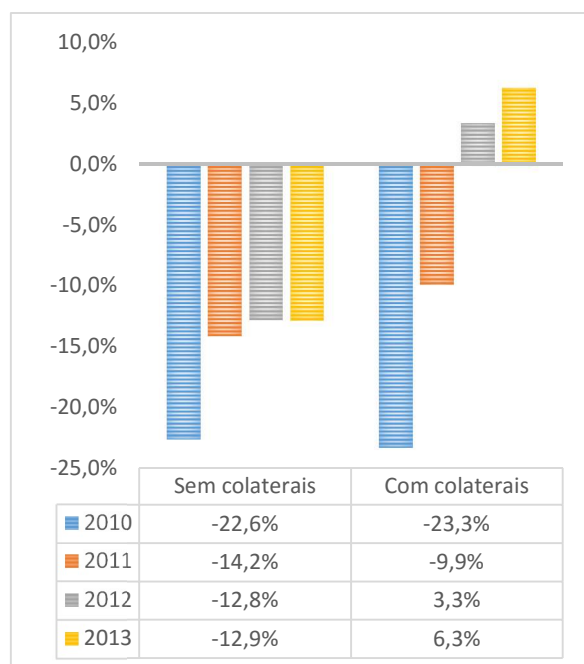
### Colaterais

**CCF Forçado ao intervalo  $[-1,1]$**

**CCF forçado ao intervalo  $[0,1]$**

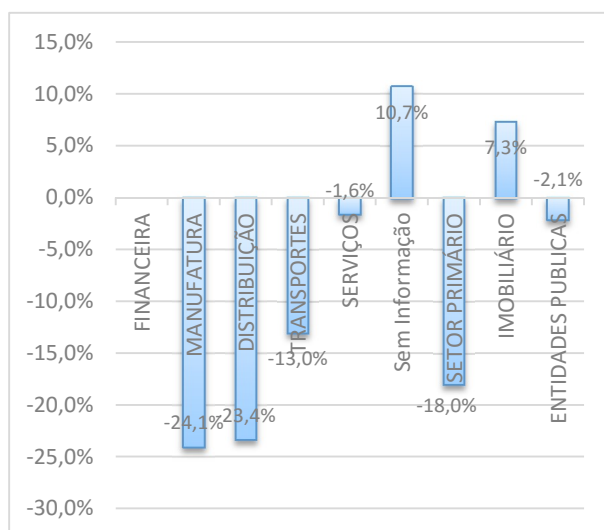


Avaliação do Impacto do Horizonte Temporal nas Estimativas de EAD de Linhas de Crédito com Limites Explícitos a PME

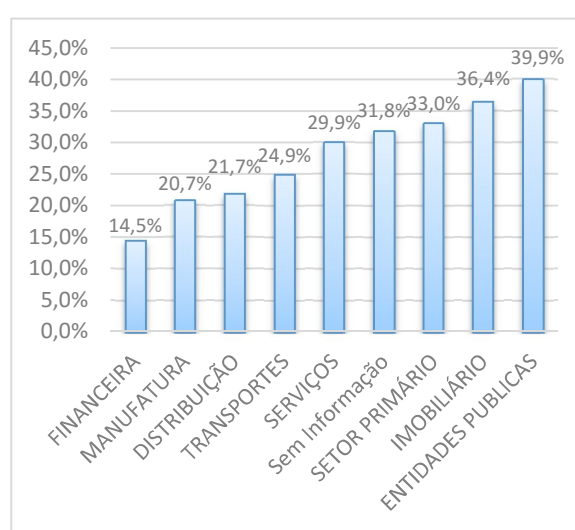


Indicador de Atividade

CCF Forçado ao intervalo  $[-1, 1]$



CCF forçado ao intervalo  $[0, 1]$

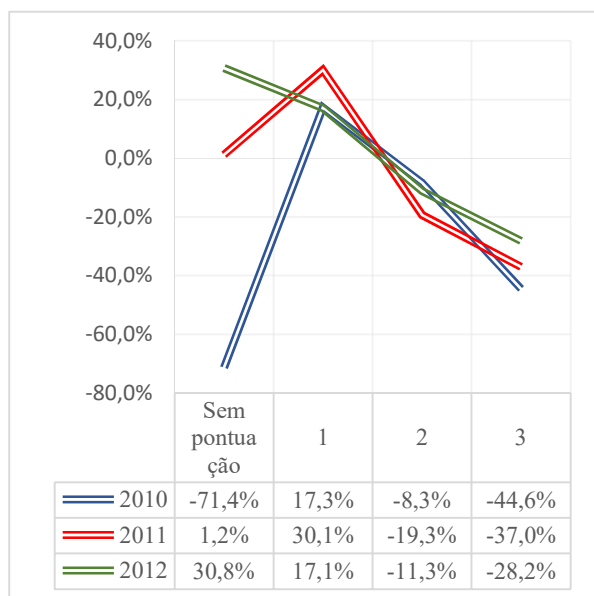




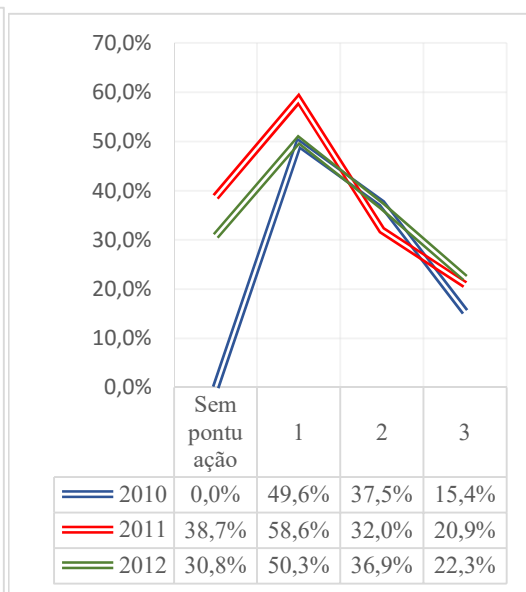
*Anexo XIII – Análise gráfica das médias do CCF forçado ao intervalo  $[-1,1]$  e do CCF forçado ao intervalo  $[0,1]$  das Variáveis discretas Pontuação, Dimensão da Empresa, Colaterais e Indicador de Atividade - Amostra de FTH restringida ao nível de utilização inferior a 99%.*

### Pontuação

CCF Forçado ao intervalo  $[-1,1]$

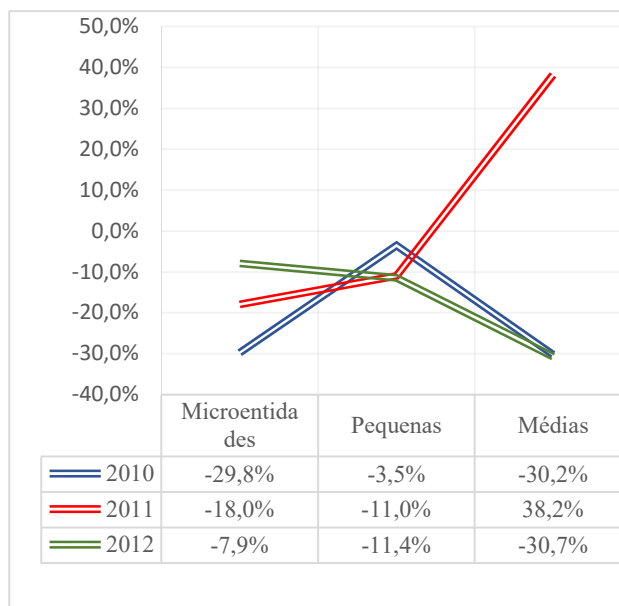


CCF forçado ao intervalo  $[0,1]$

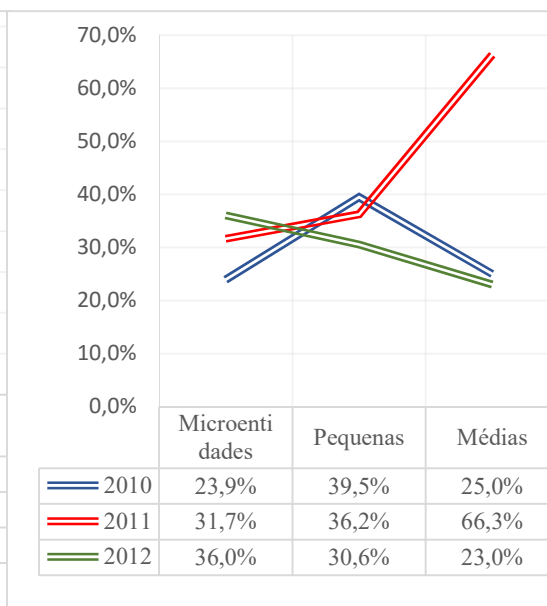


### Dimensão da Empresa

CCF Forçado ao intervalo  $[-1,1]$



CCF forçado ao intervalo  $[0,1]$

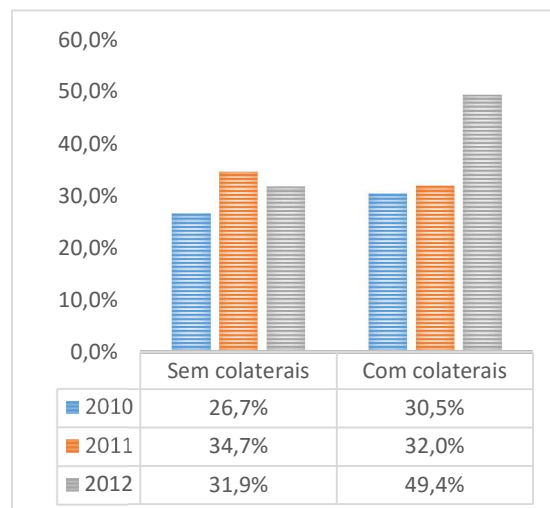
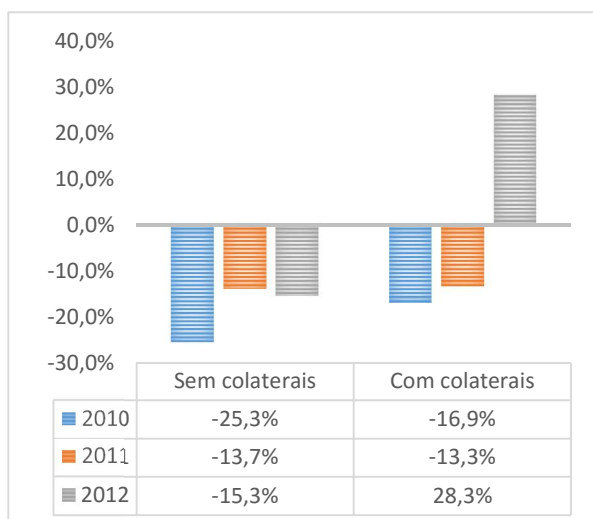


### Colaterais

CCF Forçado ao intervalo  $[-1,1]$

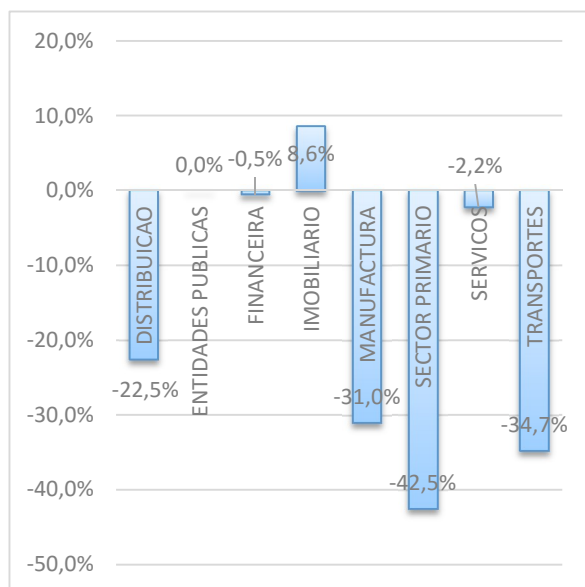
CCF forçado ao intervalo  $[0,1]$

## Avaliação do Impacto do Horizonte Temporal nas Estimativas de EAD de Linhas de Crédito com Limites Explícitos a PME

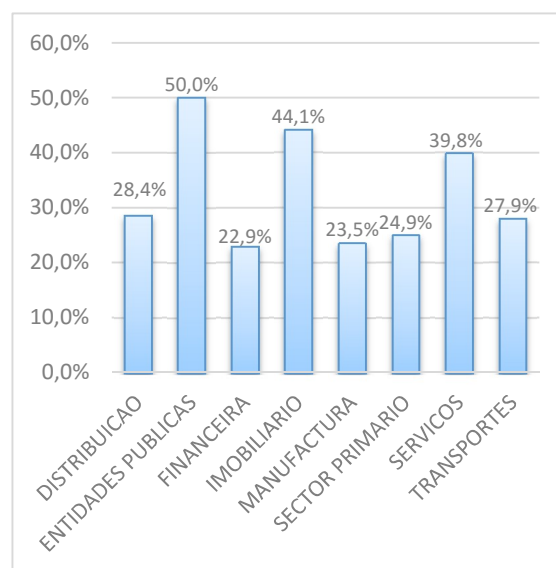


## Indicador de Atividade

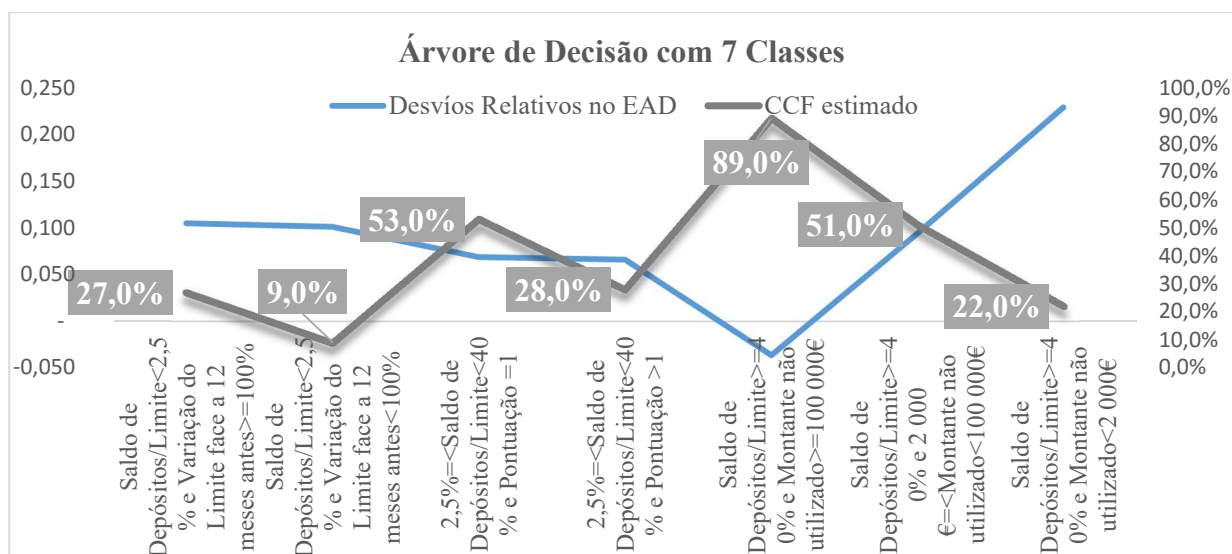
### CCF Forçado ao intervalo $[-1, 1]$



### CCF forçado ao intervalo $[0, 1]$

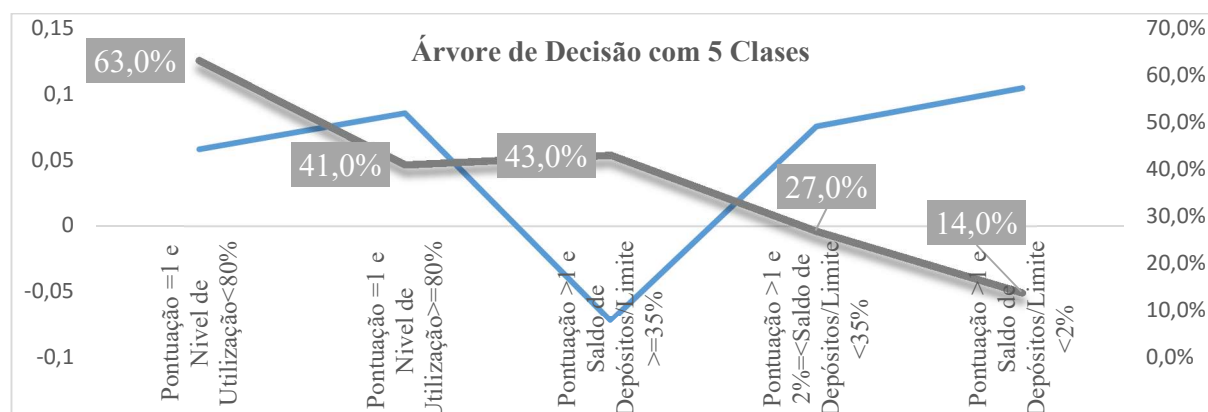


*Anexo XIV – Resultados dos Modelos de estimação baseados em médias por classes (árvores de decisão) com o CCF forçado ao intervalo unitário e a [-1,1] (“Árvore de decisão com Floor -100%”) – Amostra VTH restringida ao nível de utilização inferior a 99%.*



42

Variável	Importância <sup>43</sup>	Erro quadrático médio <sup>44</sup>	0,14536
Saldo de Depósitos/Limite	1	RMSE	0,38126
Pontuação	0,74786	Soma do quadrado dos erros	904,5896
Variação do Limite face a 12 meses antes	0,61067	Graus de liberdade totais	6223
Montante não utilizado	0,43912	% de Casos com EAD estimado inferior ao observado	28,9%



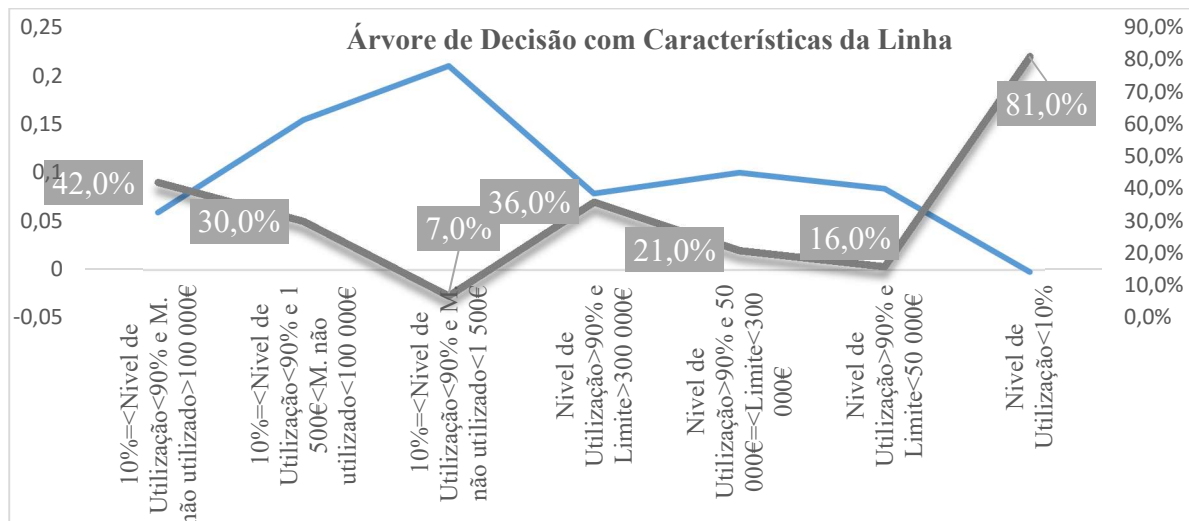
Variável	Importância	Erro quadrático médio	0,14844
----------	-------------	-----------------------	---------

<sup>42</sup> Os Desvios Relativos no EAD apresentados na linha azul representam o desvio relativo da estimativa de EAD face ao valor observado do EAD, isto é:  $\frac{EAD_t - EAD_i}{EAD_i}$

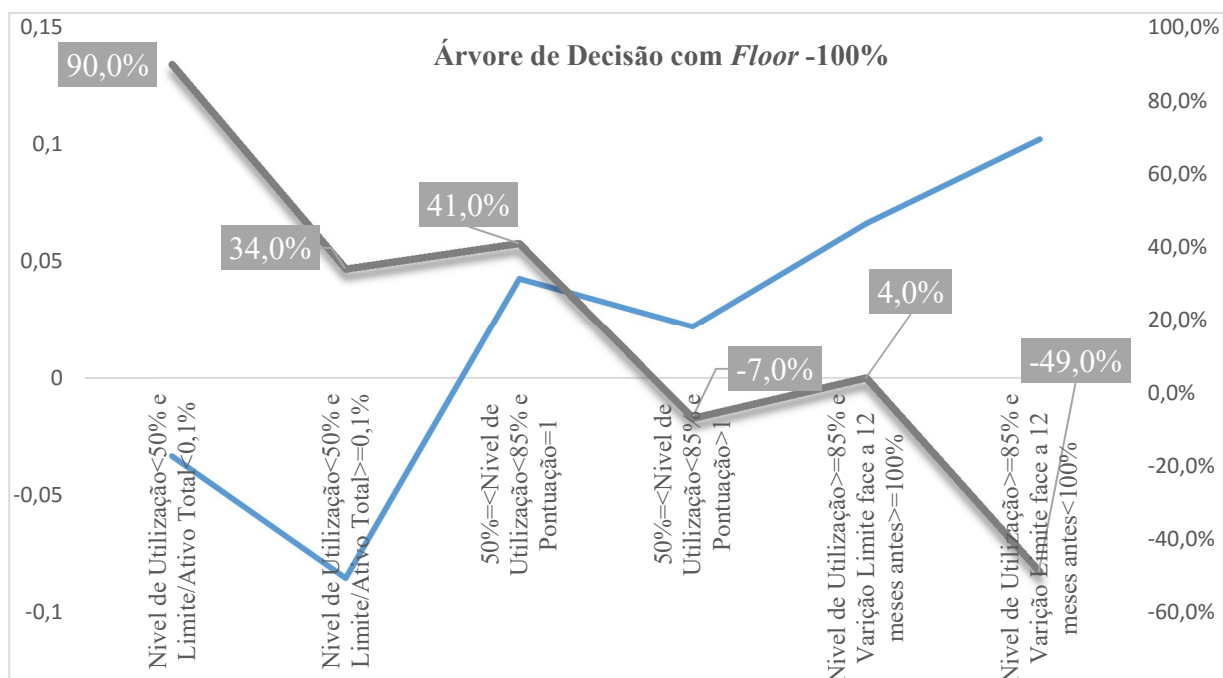
<sup>43</sup> A Importância da variável é calculada pela ferramenta SAS Enterprise Miner e indica a importância de cada variável no modelo.

<sup>44</sup> Estes valores referem-se ao CCF estimado.

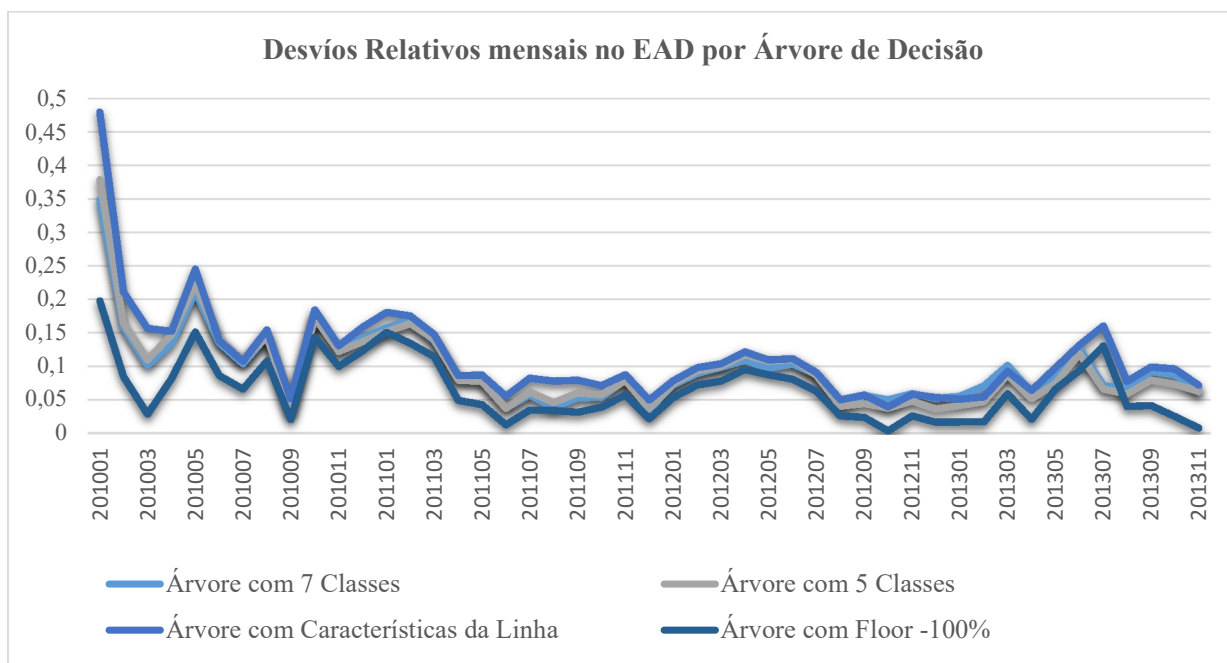
Pontuação	1	RMSE	0,3853
Saldo de Depósitos/Limite	0,71736	Soma do quadrado dos erros	923,7216
Nível de Utilização	0,41304	Graus de liberdade totais	6223
		% de Casos com EAD estimado inferior ao observado	29,1%



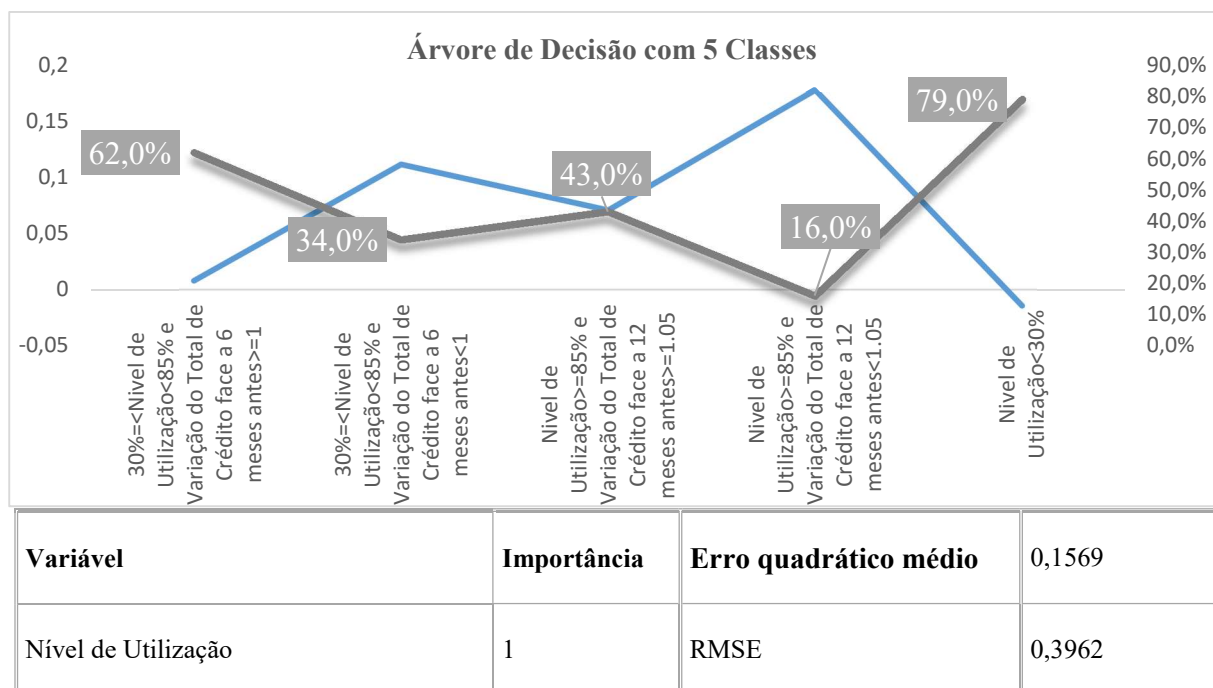
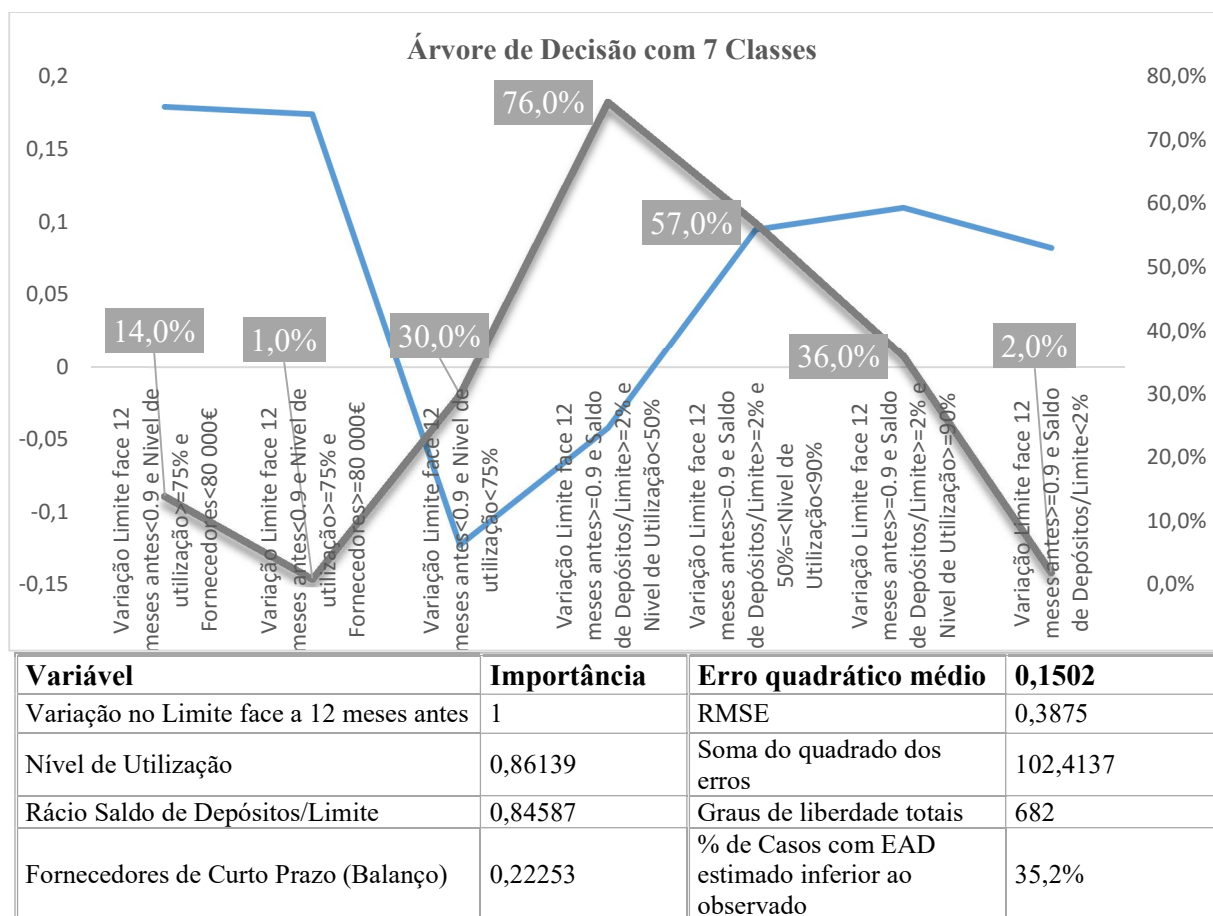
<b>Variável</b>	<b>Importância</b>	<b>Erro quadrático médio</b>	0,1530
Nível de Utilização	1	RMSE	0,3912
Limite	0,48669	Soma do quadrado dos erros	952,173
Montante não utilizado	0,4059	Graus de liberdade totais	6223
		% de Casos com EAD estimado inferior ao observado	28,8%



Variável	Importância	Erro quadrático médio	0,4725
Variação no Limite face a 12 meses antes	1	RMSE	0,6874
Nível de Utilização	0,95263	Soma do quadrado dos erros	2940,1934
Pontuação	0,42257	Graus de liberdade totais	6223
Rácio Limite/Ativo Total	0,29231	% de Casos com EAD estimado inferior ao observado	33,2%

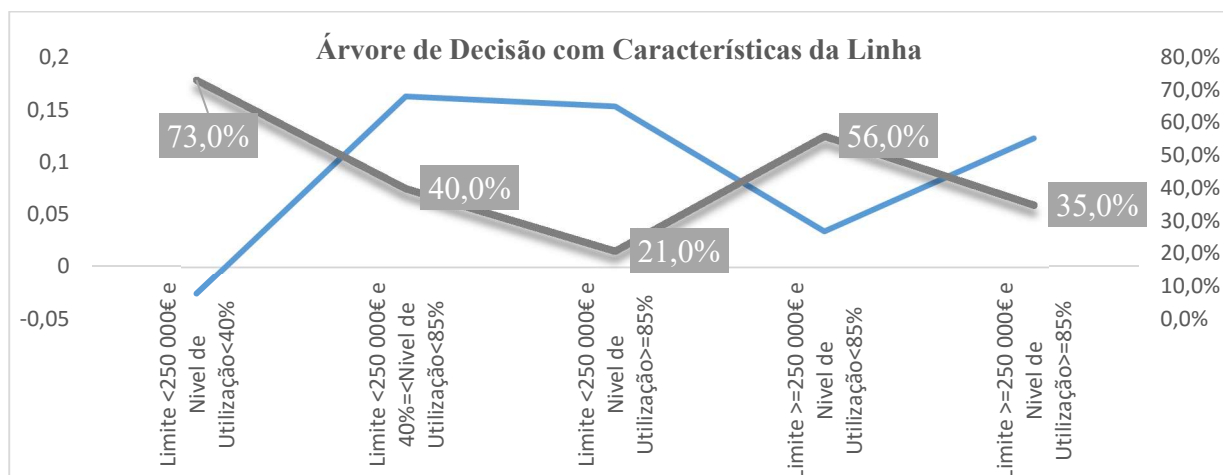


Anexo XV – Resultados dos Modelos de estimação baseados em médias por classes (árvores de decisão) com o CCF forçado ao intervalo unitário e a  $[-1,1]$  («Árvore de decisão com Floor -100%») – Amostra FTH restringida ao nível de utilização inferior a 99%.

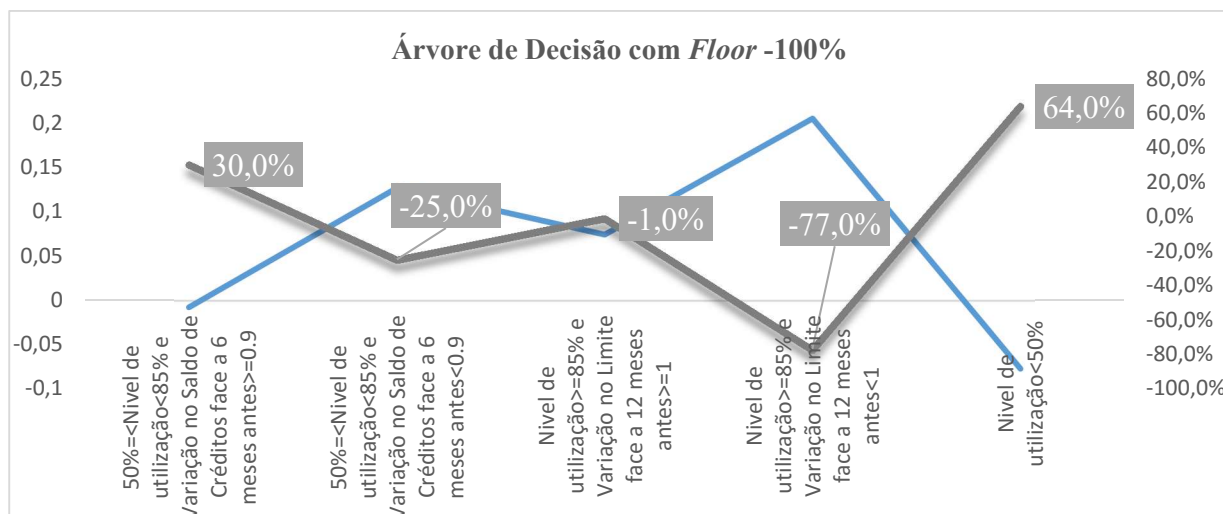


Avaliação do Impacto do Horizonte Temporal nas Estimativas de EAD de Linhas de Crédito com Limites  
Explícitos a PME

Varição saldo de créditos face a 12 meses antes	0,70698	Soma do quadrado dos erros	107,0349
Varição saldo de créditos face a 6 meses antes	0,52596	Graus de liberdade totais	682
		% de Casos com EAD estimado inferior ao observado	34,5%

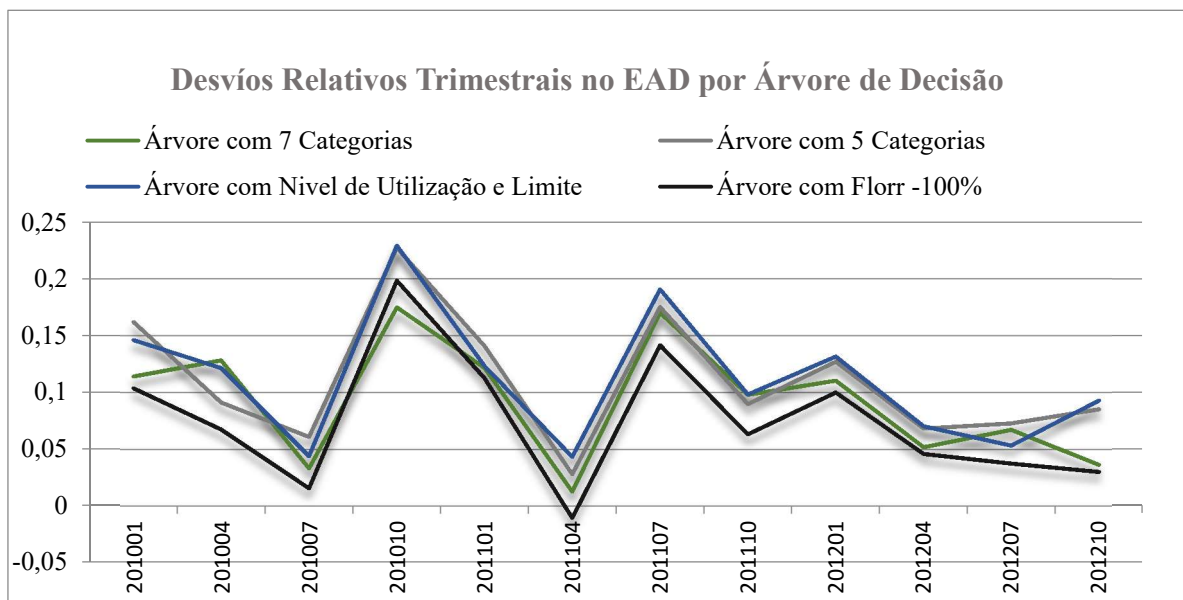


<b>Variável</b>	<b>Importância</b>	<b>Erro quadrático médio</b>	<b>0,1711</b>
Nível de Utilização	1	RMSE	0,4136
Limite	0,46609	Soma do quadrado dos erros	116,6864
		Graus de liberdade totais	682
		% de Casos com EAD estimado inferior ao observado	34,5%

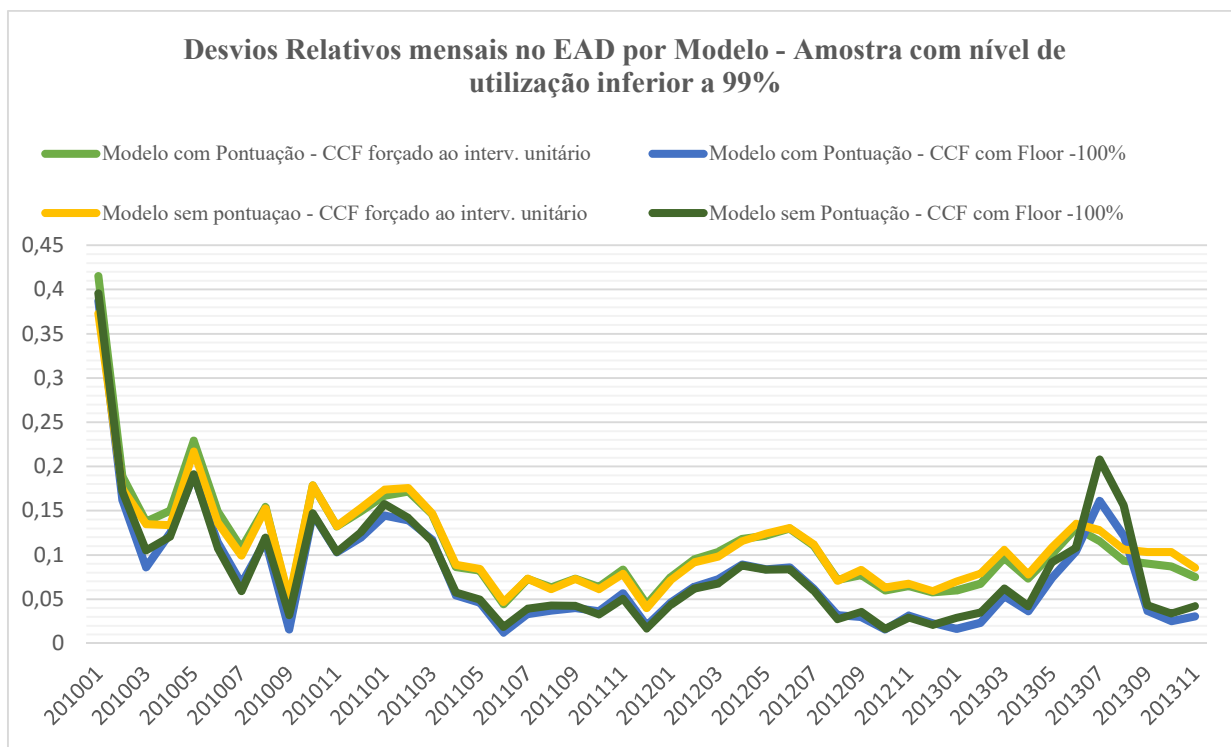


<b>Variável</b>	<b>Importância</b>	<b>Erro quadrático médio</b>	<b>0,5039</b>
Nível de Utilização	1	RMSE	0,7099

Varição no Limite face a 12 meses antes	0,90344	Soma do quadrado dos erros	343,6609
Varição no Saldo de créditos face a 6 meses antes	0,36463	Graus de liberdade totais	682
		% de Casos com EAD estimado inferior ao observado	37,4%

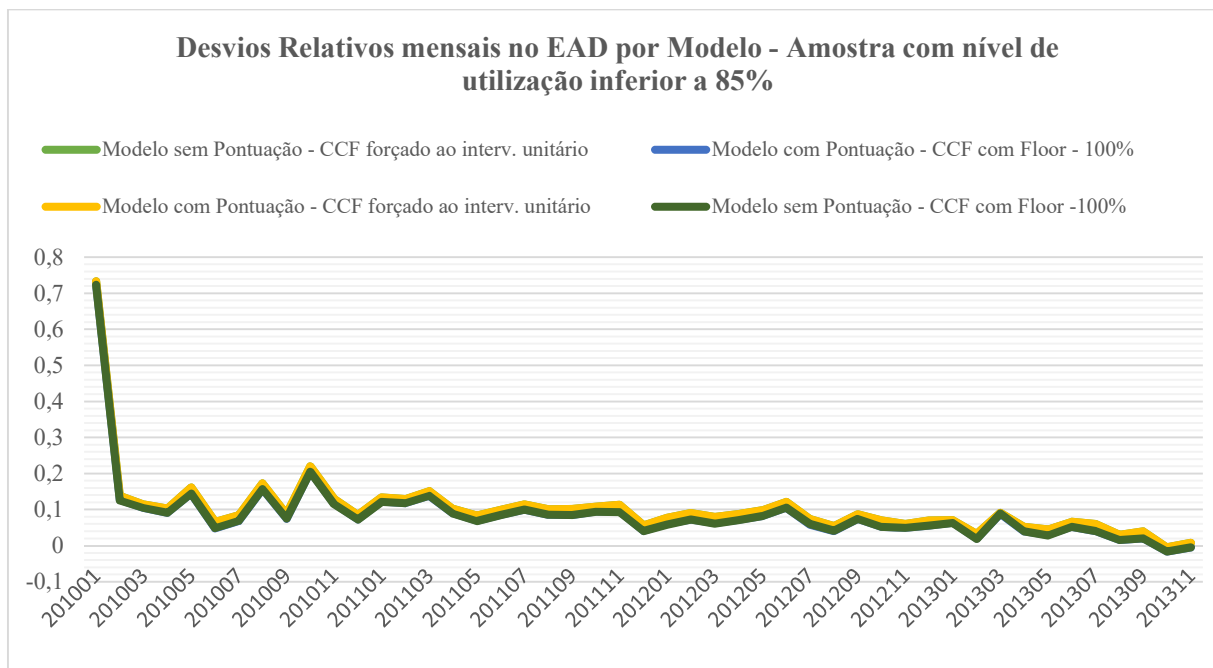


*Anexo XVI – Análise gráfica dos desvios relativos mensais no EAD dos modelos de regressão – Amostra VTH restringida ao nível de utilização inferior a 99%.*

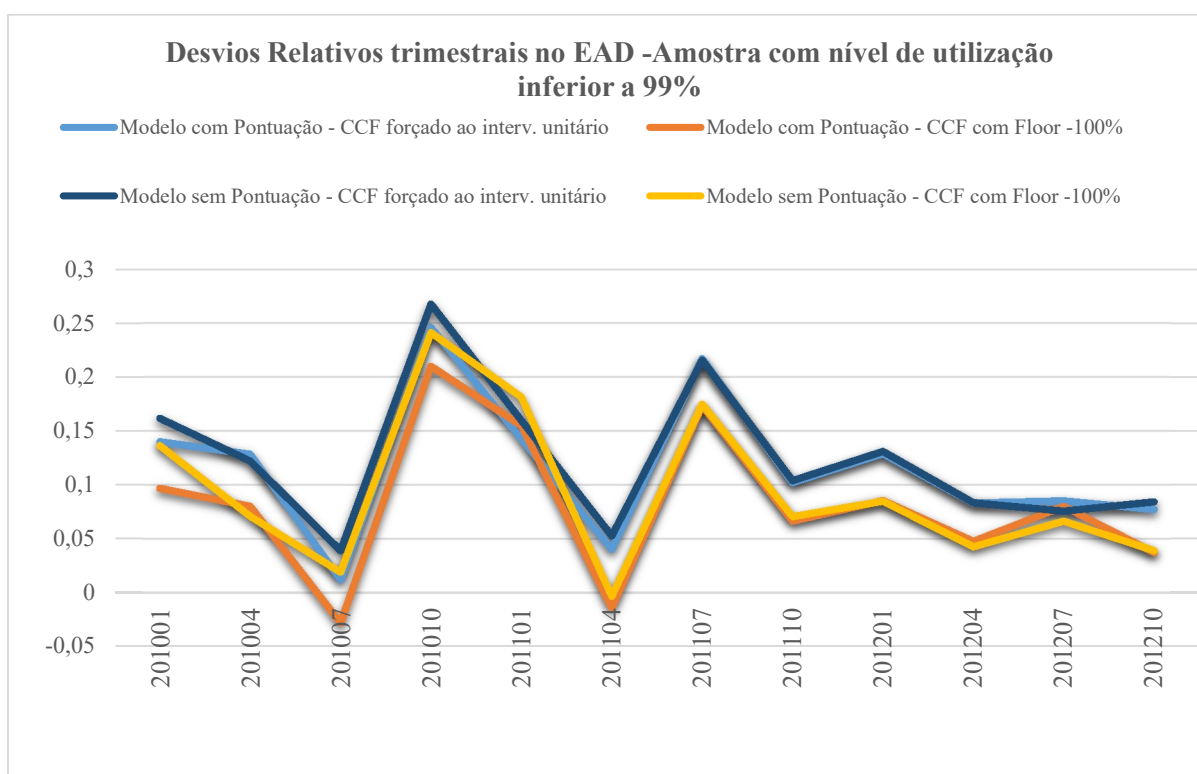




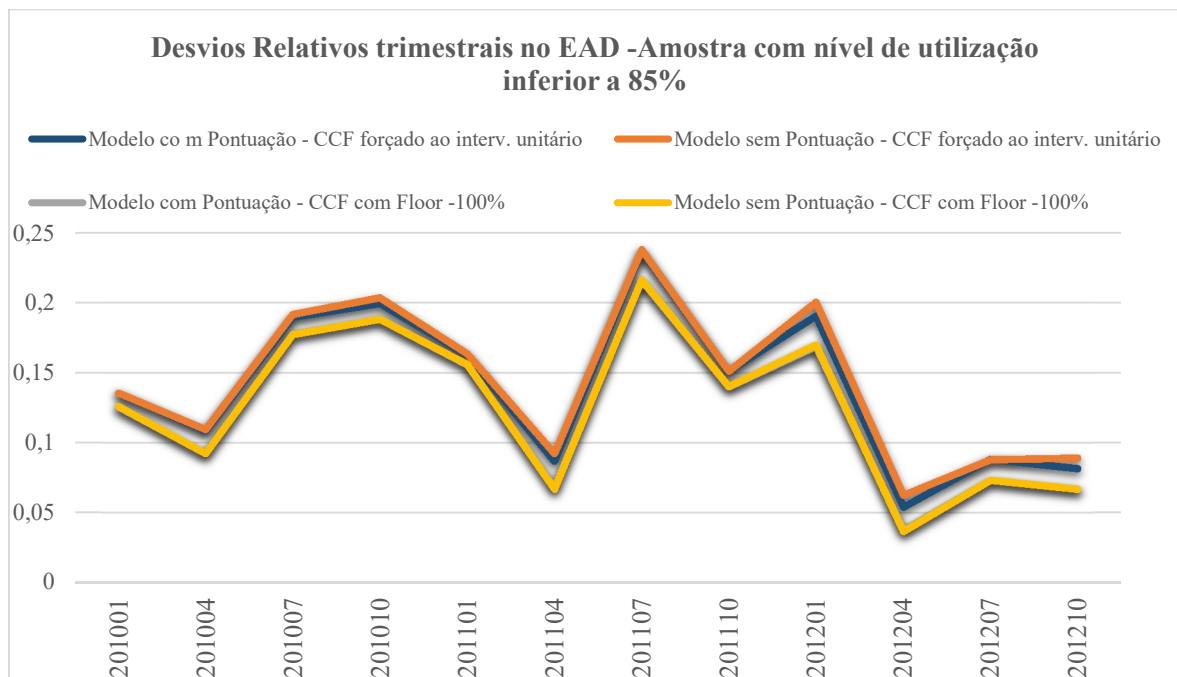
*Anexo XVII – Análise gráfica dos desvios relativos mensais no EAD dos modelos de regressão – Amostra VTH restringida ao nível de utilização inferior a 85%.*



*Anexo XVIII – Análise gráfica dos desvios relativos trimestrais no EAD dos modelos de regressão – Amostra FTH restringida ao nível de utilização inferior a 99%.*



Anexo XIX – Análise gráfica dos desvios relativos trimestrais no EAD dos modelos de regressão – Amostra FTH restringida ao nível de utilização inferior a 85%.

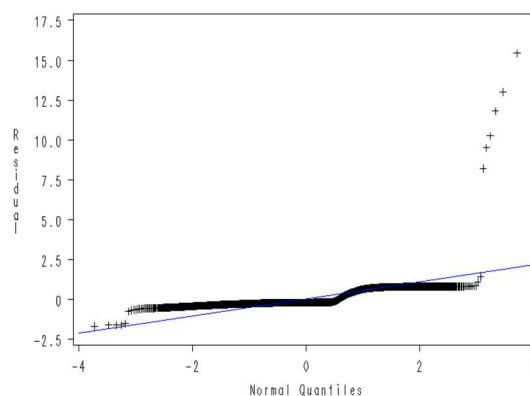
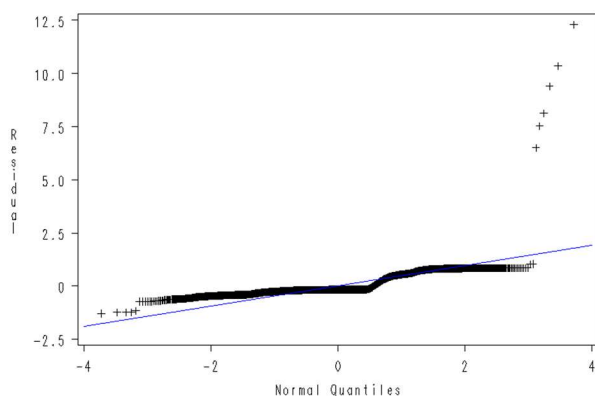


Anexo XX – Análise de resíduos dos modelos de regressão da AmostraVTH restringida ao nível de utilização inferior a 99%

### Testes de Normalidade dos Resíduos

**Modelo com Pontuação – CCF forçado ao interv. unitário**

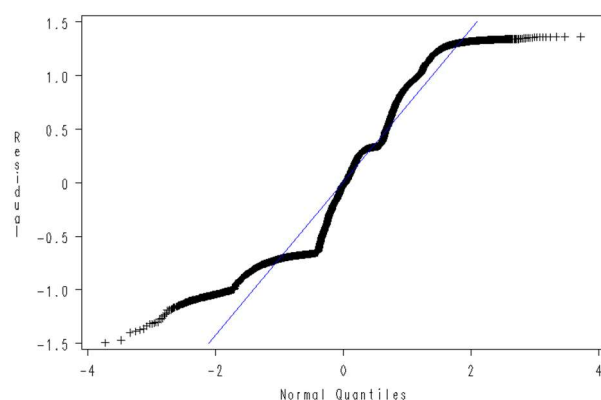
**Modelo sem Pontuação – CCF forçado ao interv. unitário**



Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.311967	Pr > D	<0.0100
Cramer-von Mises	W-Sq	114.9283	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	558.4378	Pr > A-Sq	<0.0050

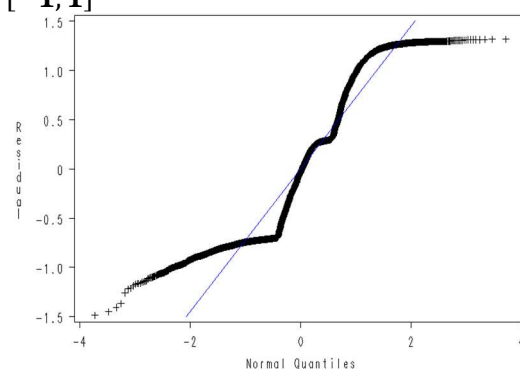
Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.312945	Pr > D	<0.0100
Cramer-von Mises	W-Sq	135.9222	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	678.5884	Pr > A-Sq	<0.0050

**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**



Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.162947	Pr > D	<0.0100
Cramer-von Mises	W-Sq	23.0546	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	154.0879	Pr > A-Sq	<0.0050

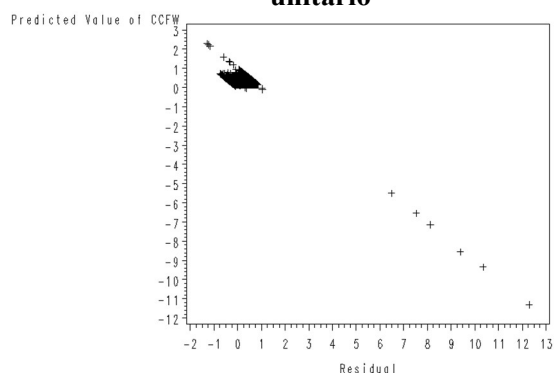
**Modelo sem Pontuação – CCF forçado a  $[-1, 1]$**



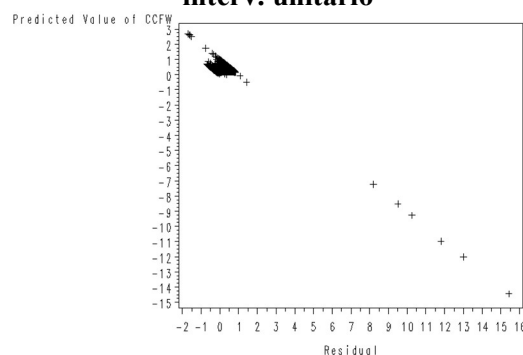
Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.169728	Pr > D	<0.0100
Cramer-von Mises	W-Sq	28.90548	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	206.7951	Pr > A-Sq	<0.0050

### Análise Valor estimados vs Resíduos

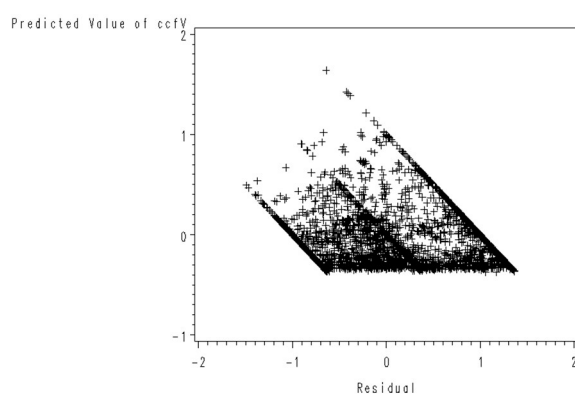
**Modelo com Pontuação – CCF forçado ao interv. unitário**



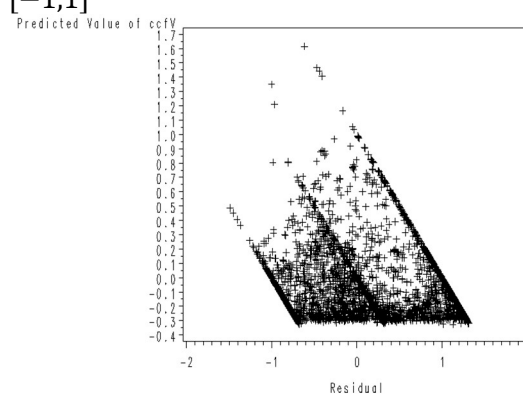
**Modelo sem Pontuação – CCF forçado ao interv. unitário**



**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**



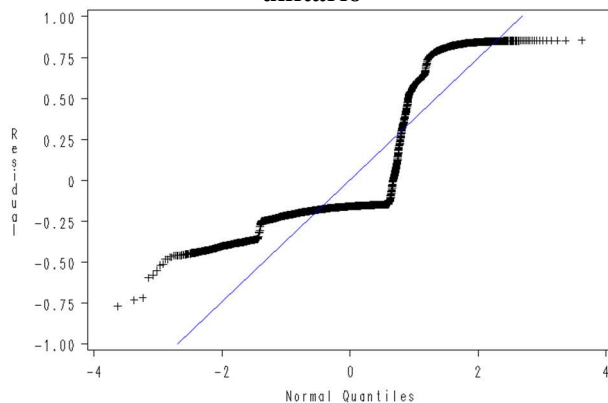
**Modelo sem Pontuação – CCF forçado a  $[-1, 1]$**



*Anexo XXI – Análise de resíduos dos modelos de regressão da Amostra VTH restringida  
ao nível de utilização inferior a 85%*

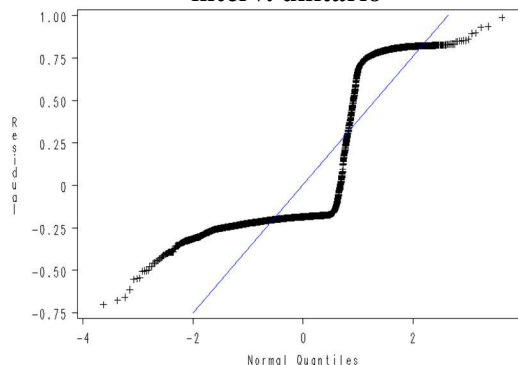
**Testes de Normalidade dos Resíduos**

**Modelo com Pontuação – CCF forçado ao interv.  
unitário**



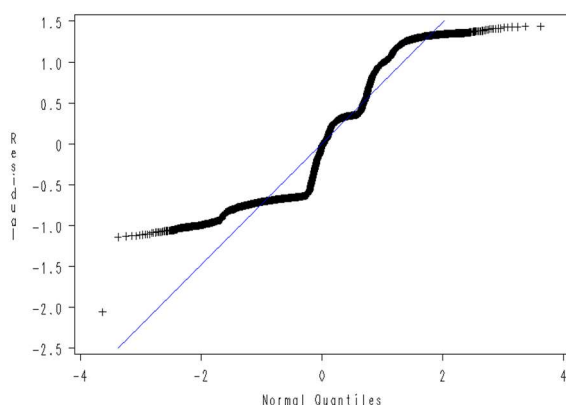
Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.37976	Pr > D	<0.0100
Cramer-von Mises	W-Sq	119.8188	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	606.9157	Pr > A-Sq	<0.0050

**Modelo sem Pontuação – CCF forçado ao  
interv. unitário**



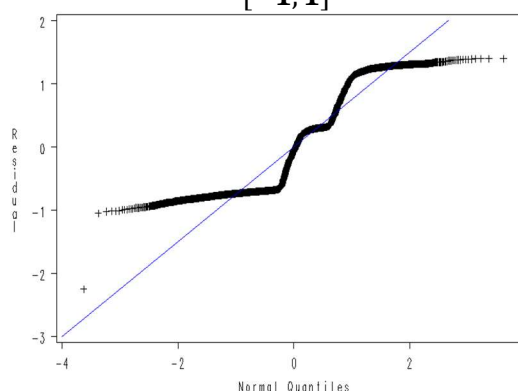
Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.376348	Pr > D	<0.0100
Cramer-von Mises	W-Sq	139.0851	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	725.2009	Pr > A-Sq	<0.0050

**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**



Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.205017	Pr > D	<0.0100
Cramer-von Mises	W-Sq	27.65755	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	176.3776	Pr > A-Sq	<0.0050

**Modelo sem Pontuação – CCF forçado a  
 $[-1, 1]$**

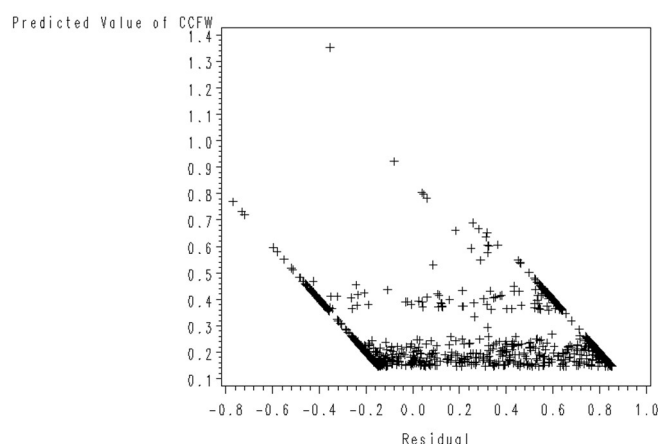


Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.211105	Pr > D	<0.0100
Cramer-von Mises	W-Sq	32.61844	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	219.6355	Pr > A-Sq	<0.0050

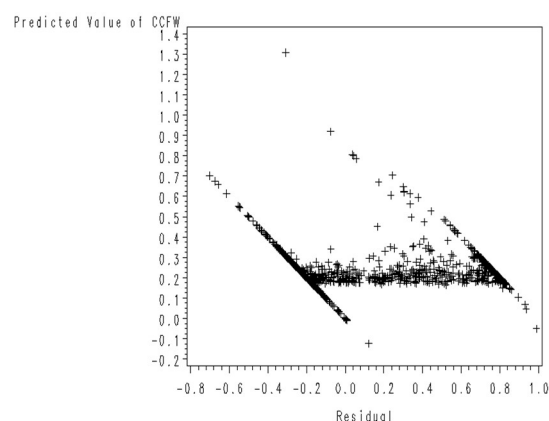
**Análise Valor estimados vs Resíduos**

**Modelo com Pontuação – CCF forçado ao interv.  
unitário**

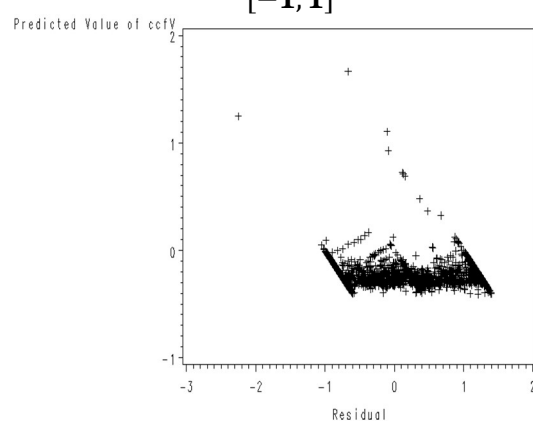
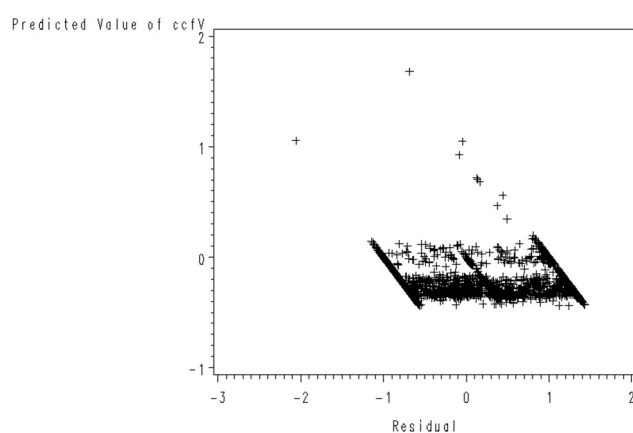
**Modelo sem Pontuação – CCF forçado ao  
interv. unitário**



**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**



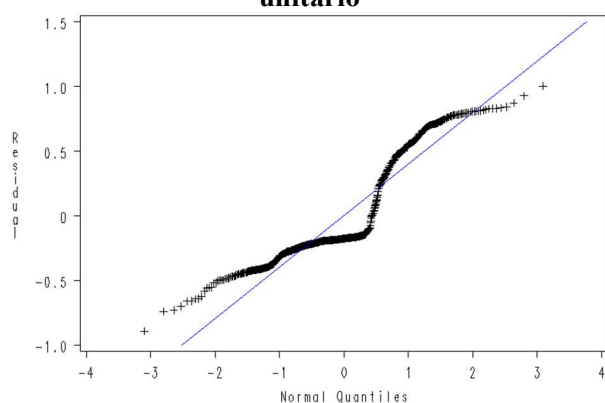
**Modelo sem Pontuação – CCF forçado a  $[-1, 1]$**



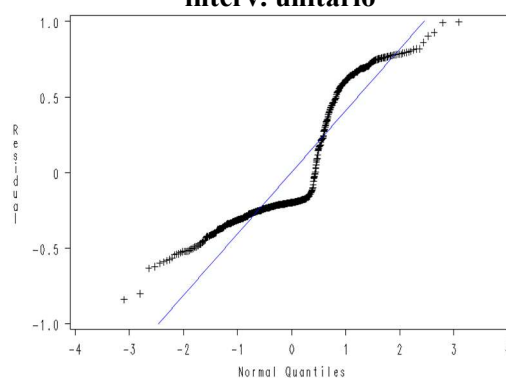
*Anexo XXII – Análise de resíduos dos modelos de regressão da Amostra de FTH restrigida ao nível de utilização inferior a 99%*

### Testes de Normalidade dos Resíduos

**Modelo com Pontuação – CCF forçado ao interv. unitário**

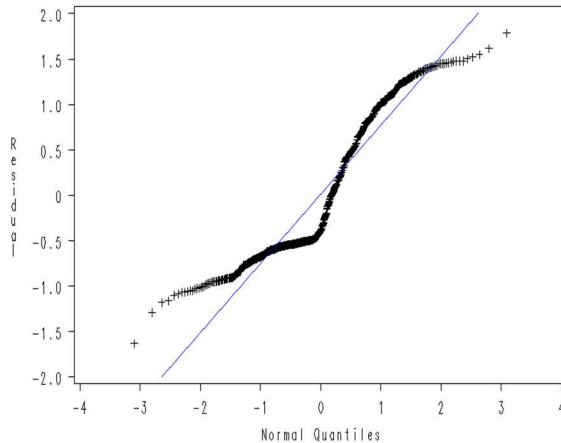


**Modelo sem Pontuação – CCF forçado ao interv. unitário**



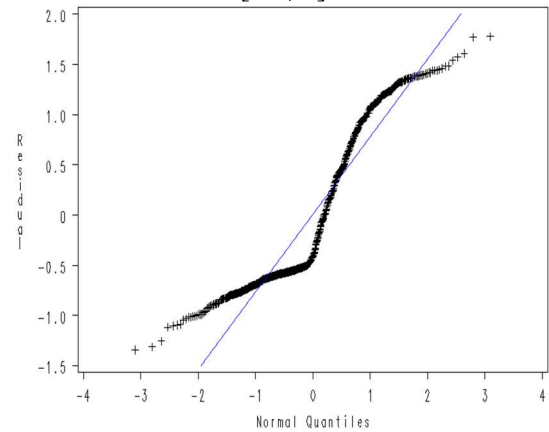
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.865202	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.270052	Pr > D	<0.0100
Cramer-von Mises	W-Sq	8.134438	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	40.7506	Pr > A-Sq	<0.0050

**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.839405	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.281942	Pr > D	<0.0100
Cramer-von Mises	W-Sq	9.402224	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	48.56634	Pr > A-Sq	<0.0050

**Modelo sem Pontuação – CCF forçado a  $[-1, 1]$**

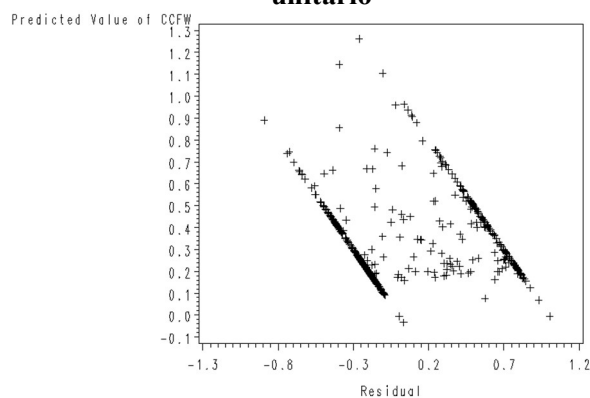


Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.900597	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.204905	Pr > D	<0.0100
Cramer-von Mises	W-Sq	4.832961	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	26.62009	Pr > A-Sq	<0.0050

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.882735	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.211565	Pr > D	<0.0100
Cramer-von Mises	W-Sq	5.528741	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	31.26784	Pr > A-Sq	<0.0050

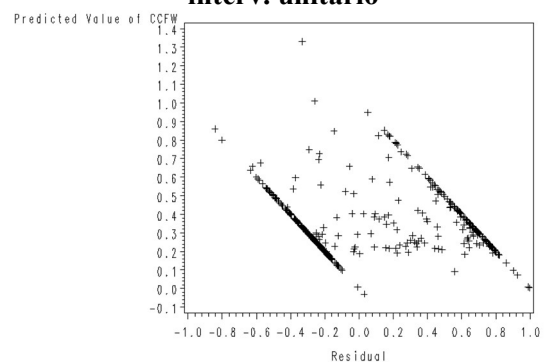
### Análise Valor estimados vs Resíduos

**Modelo com Pontuação – CCF forçado ao interv. unitário**

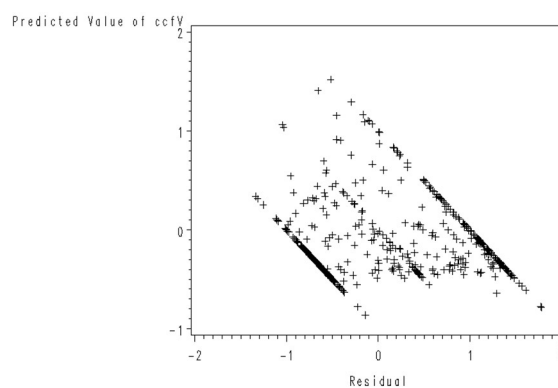
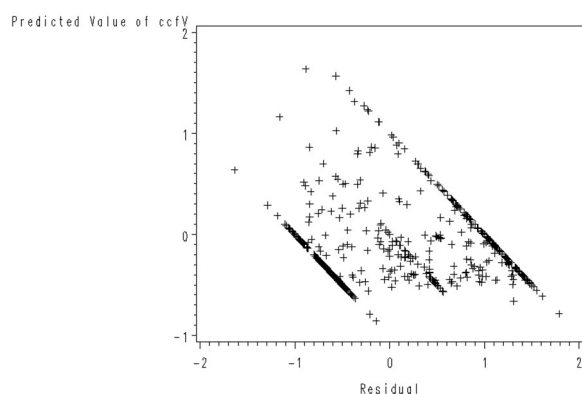


**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**

**Modelo sem Pontuação – CCF forçado ao interv. unitário**



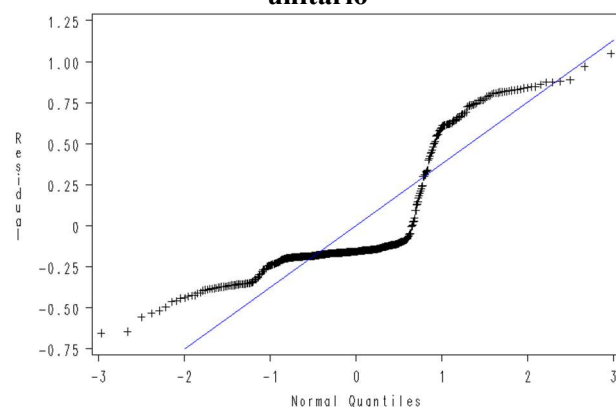
**Modelo sem Pontuação – CCF forçado a  $[-1, 1]$**



*Anexo XXIII – Análise de resíduos dos modelos de regressão da AmostraFTH restringida ao nível de utilização inferior a 85%*

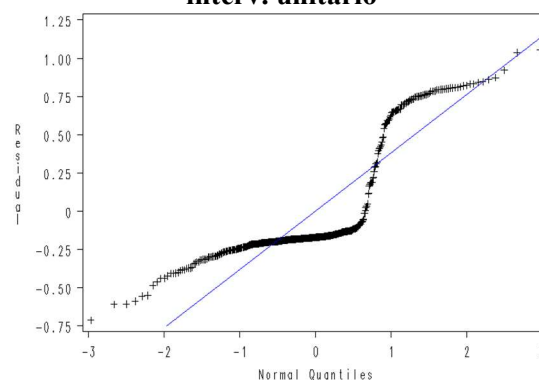
### Testes de Normalidade dos Resíduos

#### Modelo com Pontuação – CCF forçado ao interv. unitário



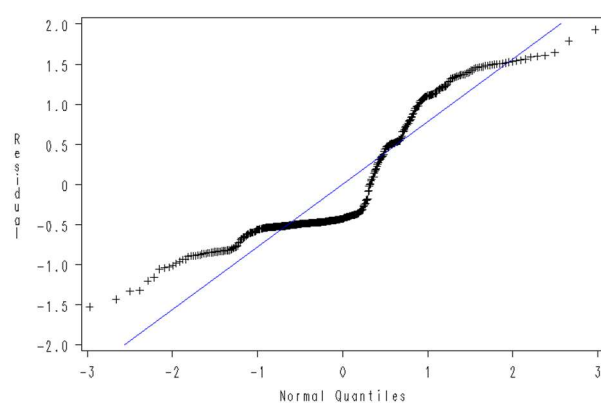
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.779484	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.311349	Pr > D	<0.0100
Cramer-von Mises	W-Sq	8.812919	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	44.53175	Pr > A-Sq	<0.0050

#### Modelo sem Pontuação – CCF forçado ao interv. unitário

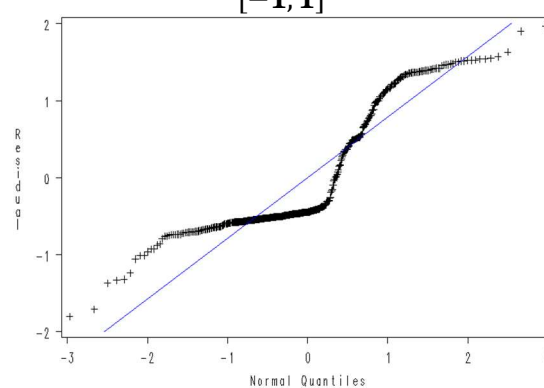


Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.753206	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.320707	Pr > D	<0.0100
Cramer-von Mises	W-Sq	9.928522	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	50.81408	Pr > A-Sq	<0.0050

#### Modelo com Pontuação – CCF forçado a $[-1, 1]$



#### Modelo sem Pontuação – CCF forçado a $[-1, 1]$

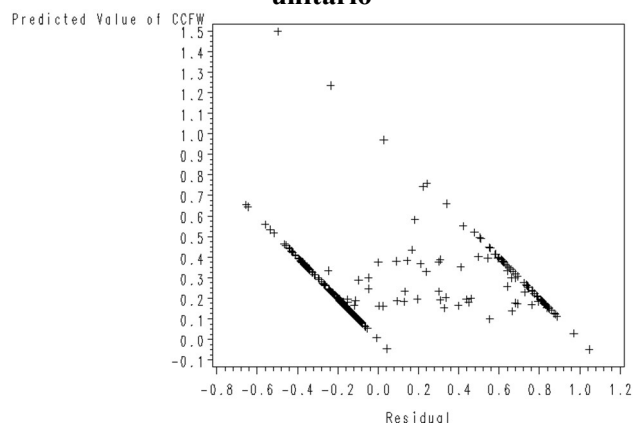


Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.864931	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.253229	Pr > D	<0.0100
Cramer-von Mises	W-Sq	5.115708	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	26.69353	Pr > A-Sq	<0.0050

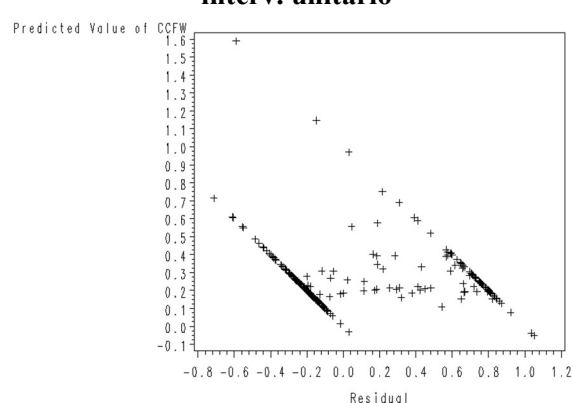
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.841949	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.263406	Pr > D	<0.0100
Cramer-von Mises	W-Sq	5.906289	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	31.54007	Pr > A-Sq	<0.0050

### Análise Valor estimados vs Resíduos

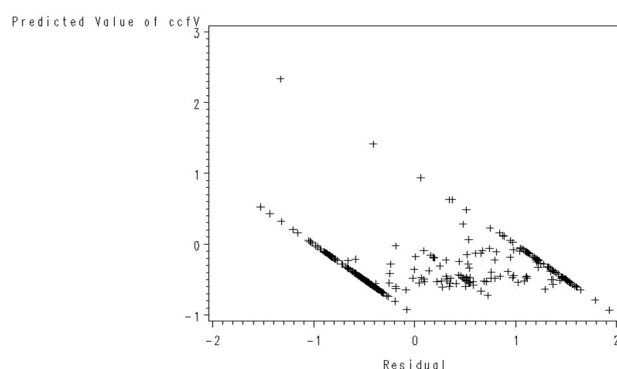
**Modelo com Pontuação – CCF forçado ao interv. unitário**



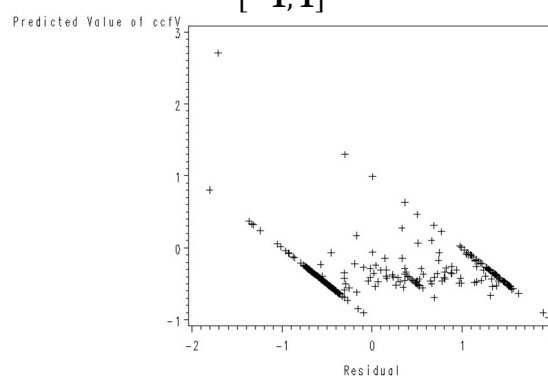
**Modelo sem Pontuação – CCF forçado ao interv. unitário**



**Modelo com Pontuação – CCF forçado a  $[-1, 1]$**

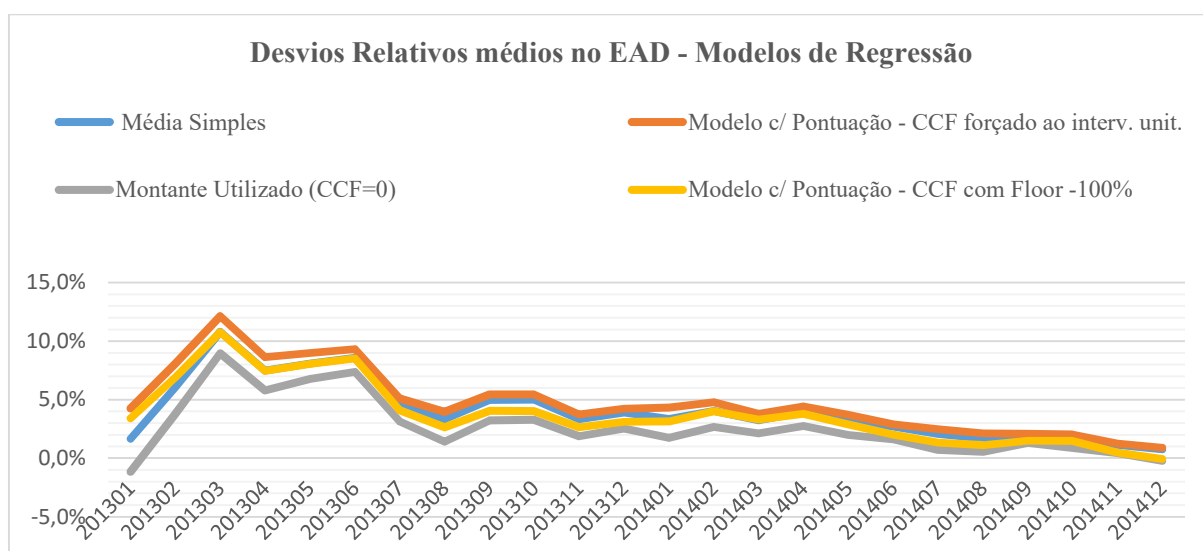
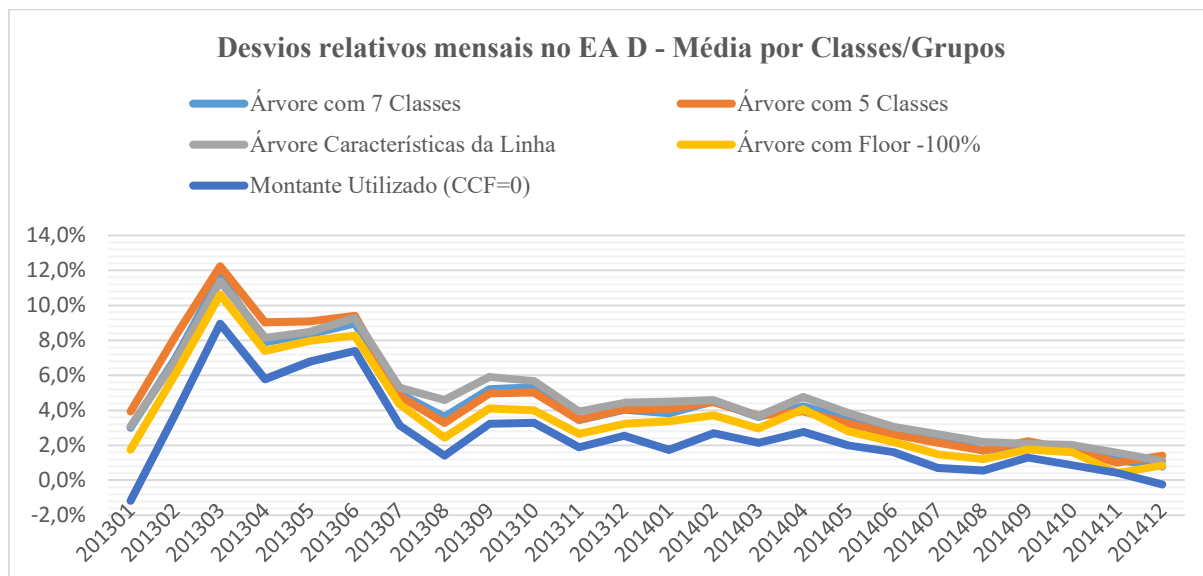


**Modelo sem Pontuação – CCF forçado a  $[-1, 1]$**





*Anexo XXIV – Desvios relativos mensais no EAD dos modelos de Horizonte Temporal*  
*Variável – Amostra de Validação*



Anexo XXV – Desvios relativos mensais no EAD dos modelos de Horizonte Temporal Fixo – Amostra de Validação

